

# Confidence intervals for the means of the selected populations

Claudio Fuentes\*

*Department of Statistics  
Oregon State University, Corvallis, OR 97331  
e-mail: [fuentes@stat.oregonstate.edu](mailto:fuentes@stat.oregonstate.edu)*

George Casella†

*Department of Statistics  
University of Florida, Gainesville, FL 32611*

and

Martin T. Wells‡

*Department of Statistical Science  
Cornell University, Ithaca, NY 14853  
e-mail: [mtw1@cornell.edu](mailto:mtw1@cornell.edu)*

**Abstract:** Consider an experiment in which  $p$  independent populations  $\pi_i$  with corresponding unknown means  $\theta_i$  are available, and suppose that for every  $1 \leq i \leq p$ , we can obtain a sample  $X_{i1}, \dots, X_{in}$  from  $\pi_i$ . In this context, researchers are sometimes interested in selecting the populations that yield the largest sample means as a result of the experiment, and then estimate the corresponding population means  $\theta_i$ . In this paper, we present a frequentist approach to the problem and discuss how to construct simultaneous confidence intervals for the means of the  $k$  selected populations, assuming that the populations  $\pi_i$  are independent and normally distributed with a common variance  $\sigma^2$ . The method, based on the minimization of the coverage probability, obtains confidence intervals that attain the nominal coverage probability for any  $p$  and  $k$ , taking into account the selection procedure.

**Keywords and phrases:** Confidence intervals, selected means, selected populations, asymmetric intervals, simultaneous inference, frequentist estimation.

Received February 2016.

## 1. Introduction

Given a set of  $p$  available features, researchers must often determine which one is the best, or simply rank them according to a certain prespecified criteria. For

---

\*To whom correspondence should be addressed.

†Distinguished Professor at the University of Florida until his passing in 2012.

‡Partially supported by NSF grant DMS-1611893 and NIH grant U19 AI111143.

instance, researchers may be interested in determining what treatment is more efficient in fighting a certain disease, or ranking the level of gene expression in a genomics experiment. This type of problems is commonly referred to as *ranking and selection procedures* and specific solutions and methods have been proposed in the literature since the second half of the 20th century, with a start that is usually traced back to the pathbreaking works of Bechhofer [2] and Gupta & Sobel [16].

In his paper, Bechhofer presents a single sample multiple decision procedure for ranking means of normal populations. Assuming the variances of the populations are known, he is able to obtain closed form expressions for the probabilities of a correct selection for a number of cases. The solution is based on the use of an *indifferent zone*, that allows him to determine a minimum guaranteed probability of selecting the population with the largest mean, as long as that mean is separated from the second largest by a prespecified distance  $\delta$  [see 3]. Alternatively, Gupta and coauthors have pioneered the subset selection approach, in which a subset of populations is selected with a minimum probability guarantee of containing the largest mean with certain probability  $P^*$  [see 15].

Note that both of these approaches are mainly concerned with the problem of correct selection of the population with the largest mean rather than estimation of the selected mean. This second problem has also been widely discussed in the literature, and in the following two sections we present a brief summary of the main findings, giving separate consideration to the point estimation and interval estimation procedures.

### 1.1. Point estimation

There are two formulations of the point estimation problem under selection, with subtle differences between them. Suppose that we have  $p$  populations, with unknown means  $\theta_i$  ( $1 \leq i \leq p$ ). Assuming that for every  $1 \leq i \leq p$  we can obtain a sample  $X_{i1}, \dots, X_{in_i}$  from the population  $\pi_i$ , we can either:

1. Attempt to select the population that has the largest parameter, *i.e.*  $\max\{\theta_1, \dots, \theta_p\}$ , and estimate its value.
2. Select the population that gives the largest sample mean, and estimate the corresponding  $\theta_i$ .

The first of these problems has been widely discussed in the literature. For example, Blumenthal & Cohen [6] consider estimating the larger mean from two normal populations and compare different estimators, but they do not discuss how to make the selection. In this direction, Guttman & Tiao [17] propose a Bayesian procedure consisting of the maximization of the expected posterior utility for a certain utility function  $U(\theta_i)$ . In the same direction, but from a frequentist perspective, Saxena & Tong [27], Saxena [26], and Chen & Dudewicz [9] consider point and interval estimation of the largest mean.

Surprisingly, despite its relevance from a practitioner's perspective, the second problem has received less attention. In this context, a common and widely

used estimator is  $\delta(\mathbf{X}) = \sum_{i=1}^p X_i I(X_i = X_{(1)})$ , where  $X_i$  is the sample mean of the  $i$ -th population,  $X_{(1)} = \max\{X_1, \dots, X_p\}$ , and  $I(\cdot)$  denotes the indicator function. Properties of this estimator have been discussed in the literature and is known to be biased [see 21]. This is particularly evident when all the populations are identically distributed (the iid case), where the problem of bias arises from the estimation of the common mean using an extreme observation.

Other properties of the estimator have also been evaluated. In particular, minimaxity and admissibility of the naive estimator  $\delta(\mathbf{X})$  were first established by Stein [29] for the case  $p = 2$ . For larger dimensions, Sackrowitz & Samuel-Cahn [25] proved the estimator is not minimax for the normal family when  $p \geq 3$ . Standard techniques to determine admissibility, following the ideas in Berger [5], Brown [7] and Lele [20], are not straightforward in this problem. However, Brown [8] is often credited to have shown admissibility of the estimator for the general case.

Nevertheless, the problem of improving the intuitive estimator is technically difficult. Dahiya [12] addresses this problem for the case of two normal populations and proposed estimators that perform better in terms of mean squared error (MSE). In this direction, progress was made by Cohen & Sackrowitz [10, 11] and Gupta & Miescke [14], where Bayes and generalized Bayes rules were obtained and studied. Venter [31] considered a bias correction approach for the problem obtaining estimators that perform well in terms of frequentist risk, and following this idea, Venter & Steel [33] introduced later the  $\omega$ -estimators which basically are a weighted average of the order statistics. Despite of these results, performance theorems are scarce. Some exception are Hwang [18], who proposes an empirical Bayes estimator and shows that it performs better in terms of the Bayes risk with respect to any normal prior, and Sackrowitz & Samuel-Cahn [24] who find UMVUE and minimax estimators of the mean of the selected population for the negative exponential distribution.

For the selection problem under sparsity, Reid & Tibshirani [23] implicitly make such assumption, with many effect sizes  $\theta_i = 0$ . They adapted theory developed in Lee et al. [19] and consider doing post-selection inference with the Lasso. Simon & Simon [28] considered adjusting the selection bias of the naive estimate from a frequentist perspective. They start by estimating the mean by maximum likelihood, and then estimate the bias in order to achieve bias reduction. To further improve on this estimate the authors suggested estimating the second order bias as well. They also compare the bias from their frequentist approach to the bias from the empirical Bayes approach in Efron [13], which turn out to be very similar. However, an advantage of the approach in [28] is that it is not limited to the Gaussian setting.

## 1.2. Interval estimation

The problem of interval estimation is equally challenging. Typically, confidence intervals are constructed in the usual way, using the standard normal distribution as a reference to attain the desired confidence level. However, these intervals

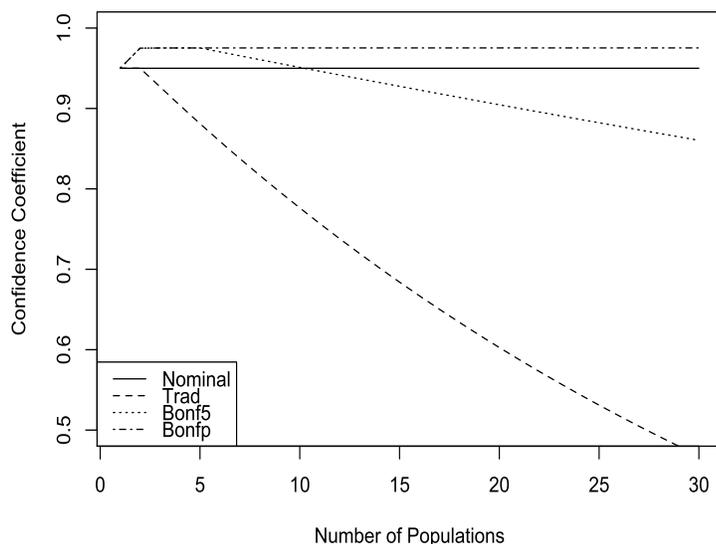


FIG 1. Confidence coefficient versus number of populations for the iid case. The solid line corresponds to the nominal level at 95%. The dashed lines correspond to the confidence coefficients for the traditional intervals and Bonferroni intervals after adjusting for  $p$  and  $k = 5$  correspondingly.

fail to maintain the nominal coverage probability as the number of populations increase. As a correction, the intervals are sometimes constructed using the Bonferroni bounds determined by the number of selected populations  $k$ . This approach is clearly a misuse of the Bonferroni correction and, in practice, the constructed intervals also fail to achieve the desired nominal level. To illustrate this issue, Figure 1 shows the results of a small numerical experiment in which we observe the behavior of the confidence coefficient against the number of populations in the iid case for the traditional intervals (Trad), Bonferroni adjusting by  $k = p$  populations (Bonfp) and Bonferroni adjusting by  $k = 5$  (Bonf5), using a nominal level of 95%. A valid approach would be to obtain simultaneous intervals for all the populations under consideration (and not just the selected ones) using the Bonferroni bounds. However, such approach is typically uninteresting either because Bonferroni intervals are known to be too conservative as the number of populations  $p$  increases, or simply because they do not offer a direct solution to the problem.

Venter [32] considers this problem and discusses how to construct confidence intervals when only one population is selected. By examination and optimization of the coverage probability, he proposes the construction of asymmetric intervals obtained as the intersection of two one-sided confidence regions. These intervals perform better than Bonferroni's simultaneous intervals in terms of length, but he does not extend his results to higher dimensions. More recently, Qiu & Hwang [22] propose an empirical Bayes approach to construct simulta-

neous confidence intervals for  $K$  selected means. Using a Bayesian framework, they consider a normal-normal model for the mean of the selected population, assuming that each population mean  $\theta_i$  follows a normal distribution. Under these assumptions, they make use of a James-Stein type estimator to center the intervals and obtain simultaneous confidence intervals that asymptotically maintain the nominal coverage probability and are substantially shorter than the ones obtained using the Bonferroni bounds. However, these intervals are not guaranteed to attain the nominal level for any  $p$  and their solution does not take in consideration the selection mechanism. Moreover, since their coverage probabilities are obtained by averaging over both sample space and prior, they do not give a valid frequentist interval.

With the explosion of big data, modern variations of this problem have become popular calling for the development of new methodologies. For instance, Benjamini & Yekutieli [4] and Zhao & Hwang [35] focus on multiplicity correction and discuss how to control for the *false coverage rate* (FCR), that is, the proportion of confidence intervals that fail to cover the parameter of interest. In their approach, given  $0 < q < 1$  they discuss how to construct confidence intervals so that the expected FCR for the selected intervals does not exceed  $q$ . In that context, the number of selected populations is a random quantity depending on  $q$ , and the resulting intervals (by construction) do not necessarily achieve the joint nominal level, which is the main purpose of this paper. We are not aware of any other attempts to solve this problem, in particular of any solutions that take explicit consideration of the selection mechanism.

In this paper, we focus on the problem of interval estimation following selection from a frequentist perspective and discuss how to construct simultaneous confidence intervals for the selected means. In Section 2 we obtain closed form expressions for the coverage probability of interest in the context and find tight lower bounds by minimizing the coverage probability function. In Section 3 we use these bounds to obtain valid confidence intervals, when the common variance  $\sigma^2$  is known or unknown, introducing asymmetric intervals. In Section 4 we present some numerical results that illustrate the performance of the proposed intervals and we finish with a brief discussion in Section 5. Some technical details are included in the Appendix.

## 2. Coverage probability results

Let  $X_1, \dots, X_p$  be independent random variables such that  $X_i \sim N(\theta_i, \sigma^2)$  where  $\theta_i$  is unknown, but the common variance  $\sigma^2$  is known. For simplicity, we take  $\sigma^2 = 1$ . Define the order statistics  $X_{(1)}, \dots, X_{(p)}$  as the sample values placed in descending order, that is, the order statistics satisfy  $X_{(1)} \geq \dots \geq X_{(p)}$ . In this context, the problem is to construct confidence intervals for the mean of the populations that give the largest sample means as a result of the experiment. Formally, for any  $0 < \alpha < 1$  and  $1 \leq k \leq p$  specified prior to the experiment, our aim is to construct simultaneous confidence intervals for  $\theta_{(1)}, \dots, \theta_{(k)}$ , based on  $X_{(1)}, \dots, X_{(k)}$ , such that the joint confidence coefficient is (at least)  $1 - \alpha$ ,

where

$$\theta_{(j)} = \sum_{i=1}^p \theta_i I(X_i = X_{(j)}),$$

for  $1 \leq j \leq k$  and  $I(\cdot)$  is the indicator function.

Even for  $k = 1$  it is not difficult to realize that the traditional confidence intervals fail to maintain the nominal coverage probability. For instance, if  $X_1, \dots, X_p \sim iid N(\theta, 1)$ , we have

$$P(\theta_{(1)} \in X_{(1)} \pm z_{\alpha/2}) = \Phi^p(z_{\alpha/2}) - \Phi^p(-z_{\alpha/2}),$$

where  $\Phi(\cdot)$  is the cumulative distribution function (cdf) of the standard normal distribution, and  $z_{\alpha/2}$  is the corresponding  $(1 - \alpha/2)$ -th percentile. Then, for  $p = 3$ , we obtain

$$\begin{aligned} P(\theta_{(1)} \in X_{(1)} \pm z_{\alpha/2}) &= \Phi^3(z_{\alpha/2}) - \Phi^3(-z_{\alpha/2}) \\ &= [\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})][\Phi^2(z_{\alpha/2}) + \Phi(z_{\alpha/2})\Phi(-z_{\alpha/2}) + \Phi^2(-z_{\alpha/2})] \\ &= [2\Phi(z_{\alpha/2}) - 1][1 - \Phi(z_{\alpha/2}) + \Phi^2(z_{\alpha/2})] \\ &< 2\Phi(z_{\alpha/2}) - 1, \end{aligned}$$

for any  $0 < \alpha < 1$ . In fact, it is easy to show that the coverage probability of the traditional intervals maintains the nominal level only for  $p = 1, 2$ , and then decreases as  $p$  approaches infinity. A similar argument shows why the Bonferroni correction based on  $k < p$  lead to the construction of invalid intervals, and constitutes a misinterpretation of the Bonferroni procedure. The problem with both of these approaches is that they fail to take into account the selection mechanism, resulting in symmetric intervals centered at  $X_{(j)}$ , and therefore ignoring the possible bias of the naive estimators. We address these issues by considering the partition of the sample space induced by the order statistics and obtaining a closed form expression for the coverage probability, which allow us to construct asymmetric intervals.

### 2.1. Selecting the best population

We begin by considering the case  $k = 1$ . The coverage probability of interest is of the form

$$P(X_{(1)} - c \leq \theta_{(1)} \leq X_{(1)} + d) = \sum_{i=1}^p P(X_i - c \leq \theta_i \leq X_i + d, X_i = X_{(1)}), \quad (1)$$

where  $c, d > 0$  are constants. Typically, we are interested in the case  $d \leq c$  to correct for the selection bias. Assuming that the variance  $\sigma^2$  is known, we define  $Z_j = (X_j - \theta_j)/\sigma$  for  $j = 1, \dots, p$ , and obtain

1.  $Z_1, \dots, Z_p$  are iid  $N(0, 1)$
2.  $X_1 \geq X_j \Leftrightarrow Z_1 - Z_j \geq \Delta_{j1}$ , with  $\Delta_{j1} = (\theta_j - \theta_1)/\sigma$ .

Hence, considering the transformation  $T : (z_1, z_2, \dots, z_p) \rightarrow (z, y_2, \dots, y_p)$  with  $z = z_1$  and  $y_j = z_1 - z_j$  ( $j = 2, \dots, p$ ), we can write the first term of the sum in (1) in terms of the parameters  $\Delta_{21}, \dots, \Delta_{p1}$ . We obtain

$$\begin{aligned} P(X_1 - c \leq \theta_1 \leq X_1 + d, X_1 = X_{(1)}) \\ &= P(-d \leq Z \leq c, Y_2 \geq \Delta_{21}, \dots, Y_p \geq \Delta_{p1}) \\ &= \frac{1}{(2\pi)^{p/2}} \int_{-d}^c \left\{ \prod_{j=2}^p \int_{\Delta_{j1}}^{\infty} e^{-\frac{1}{2}(y_j - z)^2} dy_j \right\} e^{-\frac{1}{2}z^2} dz, \end{aligned}$$

where, without loss of generality, we have taken  $\sigma^2 = 1$  to ease the notation.

Noticing that for fixed  $z$ , the integrals within the curly  $\{ \}$  brackets are essentially the tail probabilities of a normal distribution centered at  $z$ , we can write

$$P(-d \leq Z \leq c, Y_2 \geq \Delta_{21}, \dots, Y_p \geq \Delta_{p1}) = \int_{-d}^c \left\{ \prod_{j=2}^p \Phi(z - \Delta_{j1}) \right\} \phi(z) dz,$$

where  $\phi(\cdot)$  denotes the probability density function (pdf) of the standard normal distribution. It follows that we can entirely describe the probability  $P(X_{(1)} - c \leq \theta_{(1)} \leq X_{(1)} + d)$  in terms of a new set of parameters  $\Delta_{ij}$ , where  $\Delta_{ij} = \theta_i - \theta_j$  ( $1 \leq i, j \leq p$ ) satisfy

1.  $\Delta_{jj} = 0$ , for every  $j$ .
2.  $\Delta_{ij} = -\Delta_{ji}$ , for every  $i, j$ .
3. For  $j > k$ ,  $\Delta_{jk} = \Delta_{j,j-1} + \Delta_{j-1,j-2} + \dots + \Delta_{k+1,k}$ .

Under this representation, the values of the coverage probability are completely determined by the relative distances between the population means  $\theta_i$  ( $1 \leq i \leq p$ ) and therefore, we can think of the coverage probability as a function  $h = h(\mathbf{\Delta}|c, d)$ , where  $\mathbf{\Delta}$  denotes the vector of possible configurations of the  $\Delta_{ij}$ .

For instance, when  $p = 3$ , we can fully describe the problem in terms of the parameters  $\Delta_{21}$  and  $\Delta_{32}$ . Figure 2 shows the form of the coverage probability surface in terms of these parameters for two different values of  $c, d$ . If we further assume, without any loss of generality, that  $\theta_1 \leq \theta_2 \leq \theta_3$ , we can write the probability of interest as

$$\begin{aligned} P(X_{(1)} - c \leq \theta_{(1)} \leq X_{(1)} + d) &= \int_{-d}^c \Phi(z - \Delta_{21}) \Phi(z - \Delta_{32} - \Delta_{21}) \phi(z) dz \quad (2) \\ &\quad + \int_{-d}^c \Phi(z + \Delta_{21}) \Phi(z - \Delta_{32}) \phi(z) dz \\ &\quad + \int_{-d}^c \Phi(z + \Delta_{32}) \Phi(z + \Delta_{32} + \Delta_{21}) \phi(z) dz \\ &= h(\Delta_{21}, \Delta_{32}|c, d), \end{aligned}$$

where  $\Delta_{21}, \Delta_{32} \geq 0$ . Observe that a trivial lower bound for the probability in (2) is given by  $\Phi(c) - \Phi(-d)$ , obtained by conveniently setting the values of  $\Delta_{ij}$  equal to 0 or  $\infty$  in each term of the expression. In order to obtain a sharper

bound we look at the behavior of the coverage probability in order to find its minimum. Looking at the partial derivatives, we obtain

$$\begin{aligned} \frac{\partial h(\Delta_{21}, \Delta_{32} | c, d)}{\partial \Delta_{21}} &= \int_{-d}^c \Phi(z + \Delta_{32}) \phi(\Delta_{32} + \Delta_{21} + z) \phi(z) dz \\ &\quad - \int_{-d}^c \Phi(z - \Delta_{21}) \phi(\Delta_{32} + \Delta_{21} - z) \phi(z) dz \\ &\quad + \int_{-d}^c \Phi(z - \Delta_{32}) \phi(\Delta_{21} + z) \phi(z) dz \\ &\quad - \int_{-d}^c \Phi(z - \Delta_{32} - \Delta_{21}) \phi(\Delta_{21} - z) \phi(z) dz. \end{aligned} \quad (3)$$

Combining terms and changing variables, we can rewrite (3) as  $\partial h / \partial \Delta_{21} = D_1 + D_2$ , where

$$\begin{aligned} D_1 &= \left\{ \int_{\Delta_{32} + \Delta_{21} - d}^{\Delta_{32} + \Delta_{21} + c} - \int_{-d}^c \right\} \Phi(z - \Delta_{21}) \phi(\Delta_{32} + \Delta_{21} - z) \phi(z) dz, \\ D_2 &= \left\{ \int_{\Delta_{21} - d}^{\Delta_{21} + c} - \int_{-d}^c \right\} \Phi(z - \Delta_{32} - \Delta_{21}) \phi(\Delta_{21} - z) \phi(z) dz. \end{aligned}$$

The values of  $c, d$  can be chosen so that  $D_1$  and  $D_2$  are both positive (see Lemma A.1 in the Appendix), and therefore,  $\partial h / \partial \Delta_{21} > 0$ . A similar argument gives  $\partial h / \partial \Delta_{32} > 0$ , and we obtain that the coverage probability is minimized at  $\Delta_{21} = \Delta_{32} = 0$ , or equivalently, at  $\theta_1 = \theta_2 = \theta_3$ . The general result follows from induction on  $p$ , leading to the following theorem:

**Theorem 1.** *Let  $X_1, \dots, X_p$  be independent random variables with  $X_i \sim N(\theta_i, \sigma^2)$  for  $i = 1, \dots, p$ . Then, for any  $c > 0$  there exists  $0 < d_0 \leq c$  such that:*

i) If  $d_0 < d$ ,

$$\min_{\theta_1, \dots, \theta_p} P(\theta_{(1)} \in (X_{(1)} - c, X_{(1)} + d)) = \Phi^p(c) - \Phi^p(-d). \quad (4)$$

ii) If  $0 < d < d_0$ ,

$$\min_{\theta_1, \dots, \theta_p} P(\theta_{(1)} \in (X_{(1)} - c, X_{(1)} + d)) \geq \Phi^q(c) - \Phi^q(-d), \quad (5)$$

for some  $1 \leq q \leq p$ .

## 2.2. Selecting the top $k$ populations

For general  $k$ , the coverage probability of interest is of the form

$$\begin{aligned} &P(\theta_{(1)} \in (X_{(1)} - c, X_{(1)} + d), \dots, \theta_{(k)} \in (X_{(k)} - c, X_{(k)} + d)) \\ &= \sum_{j=1}^{\binom{p}{k}} P[\theta_{j_1} \in (X_{j_1} - c, X_{j_1} + d), \dots, \theta_{j_k} \in (X_{j_k} - c, X_{j_k} + d), \\ &\quad \min\{X_{j_1}, \dots, X_{j_k}\} \geq \max\{X_{j_{k+1}}, \dots, X_{j_p}\}]. \end{aligned} \quad (6)$$

Observe that each term of the sum in (6) is determined by the joint distribution of  $X_1, \dots, X_p$ . For instance, when  $(j_1, \dots, j_k) = (1, \dots, k)$ , the corresponding piece of relevant probability is

$$\begin{aligned} & P(\theta_1 \in (X_1 - c, X_1 + d), \dots, \theta_k \in (X_k - c, X_k + d), \min\{X_1, \dots, X_k\} \geq \max\{X_{k+1}, \dots, X_p\}) \\ &= \int_{\theta_1 - d}^{\theta_1 + c} \cdots \int_{\theta_k - d}^{\theta_k + c} \prod_{j=k+1}^p P_{\theta_j}(X_j \leq \min\{x_1, \dots, x_k\}) f(x_1, \dots, x_k) dx_1 \cdots dx_k \\ &= \int_{-d}^c \cdots \int_{-d}^c \prod_{j=k+1}^p \Phi(\min\{z_1 + \theta_1, \dots, z_k + \theta_k\} - \theta_j) \prod_{i=1}^k \phi(z_i) dz_i, \end{aligned} \quad (7)$$

where  $z_i = x_i - \theta_i$  for  $i = 1, \dots, k$ .

Combining (6) and (7), it is not difficult to obtain the following expression for the coverage probability

$$\begin{aligned} & P(\theta_{(1)} \in (X_{(1)} - c, X_{(1)} + d), \dots, \theta_{(k)} \in (X_{(k)} - c, X_{(k)} + d)) \\ &= \sum_{j=1}^{\binom{p}{k}} \int_{-d}^c \cdots \int_{-d}^c \prod_{m \in I_j^c} \Phi\left(\min_{\ell \in I_j} \{z_\ell + \theta_\ell\} - \theta_m\right) \prod_{\ell \in I_j} \phi(z_\ell) dz_\ell, \end{aligned} \quad (8)$$

where  $I_j = \{j_1, \dots, j_k\}$  and  $I_j^c = \{j_{k+1}, \dots, j_p\}$  are respectively the set of indices for the top  $k$  variables and the bottom  $p - k$  variables in the  $j$ -th arrangement.

Before we move forward, let us take a closer look at the formula in (8) and consider the case  $p = 6$  and  $k = 3$ . Then, the sum will have  $\binom{6}{3} = 20$  terms determined by the configurations

$$\begin{array}{cccc} 123|456 & 234|156 & 345|126 & 456|123 \\ 124|356 & 235|146 & 346|125 & \\ 125|346 & 236|145 & 356|124 & \\ 126|345 & 245|136 & & \\ 134|256 & 246|136 & & \\ 135|246 & 256|134 & & \\ 136|245 & & & \\ 145|236 & & & \\ 146|235 & & & \\ 156|234 & & & \end{array}$$

where the numbers to the left of the vertical line correspond to the indices of the set  $I_j$  (the populations being selected in the respective arrangement) and the numbers on the right to the indices of the set  $I_j^c$  (the non-selected populations). Observe that all the indices appear on the left side (and on the right side) the same number of times.

Now, suppose that  $\theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4 \leq \theta_5 \leq \theta_6$  and let  $\theta_6 \uparrow \infty$ . Then, all the terms where population 6 is on the right side approach to zero. On the other hand, for those terms where 6 appears on the left, the value of

$\Phi(\min_{\ell=1,\dots,k}\{z_\ell + \theta_\ell\} - \theta_m)$  becomes independent of  $\theta_6$ . It follows that, as  $\theta_6 \uparrow \infty$ , the coverage probability is determined by the configurations

$$\begin{array}{cccc} 12|345 & 23|145 & 34|125 & 45|123 \\ 13|245 & 24|135 & 35|124 & \\ 14|235 & 25|134 & & \\ 15|234 & & & \end{array}$$

which correspond to the number of ways to choose 2 out of 5 populations.

Repeating the argument, but letting  $\theta_5 \uparrow \infty$ , we obtain the configurations

$$\begin{array}{cc} 1|234 & 3|124 \\ 2|134 & 4|123 \end{array}$$

which are the possible ways to choose 1 out of 4 populations and we know, from Section 2, that the minimum is reached at  $\theta_1 = \theta_2 = \theta_3 = \theta_4$ .

To extend this argument for the general case ( $1 \leq k < p$ ), we first notice that the number of possible configurations is

$$\binom{p}{k} = \binom{p-1}{k} + \binom{p-1}{k-1}$$

where  $\binom{p-1}{k}$  is the number of times that any given index  $i$  appears on the right side (population  $i$  is not selected) and  $\binom{p-1}{k-1}$  is the number of arrangements that have the index  $i$  on the left side (population  $i$  is selected).

Due to the symmetry of the problem, we assume (without any loss of generality) that  $\theta_1 \leq \dots \leq \theta_p$ . Also, for every  $j$  such that  $p \in I_j$  we define

$$\begin{aligned} \mathcal{I}_j(\theta_p) &= I\left(\min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\} \geq z_p + \theta_p\right) \\ \mathcal{I}_j^c(\theta_p) &= I\left(\min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\} < z_p + \theta_p\right) \end{aligned}$$

where  $I(\cdot)$  is the indicator function. Then, we can write

$$\min_{\ell \in I_j} \{z_\ell + \theta_\ell\} = (z_p + \theta_p)\mathcal{I}_j(\theta_p) + \min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\}\mathcal{I}_j^c(\theta_p),$$

and therefore

$$\lim_{\theta_p \uparrow \infty} \prod_{m \in I_j^c} \Phi\left(\min_{\ell \in I_j} \{z_\ell + \theta_\ell\} - \theta_m\right) = \prod_{m \in I_j^c} \Phi\left(\min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\} - \theta_m\right).$$

At the same time, for every  $j$  such that  $p \in I_j^c$ , we have

$$\prod_{m \in I_j^c} \Phi\left(\min_{\ell \in I_j} \{z_\ell + \theta_\ell\} - \theta_m\right) \rightarrow 0$$

and consequently, as  $\theta_p \uparrow \infty$ , the coverage probability converges to

$$\sum_{j=1}^{\binom{p-1}{k-1}} \int_{-d}^c \cdots \int_{-d}^c \prod_{m \in I_j^c} \Phi \left( \min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\} - \theta_m \right) \prod_{\ell \in I_j} \phi(z_\ell) dz_\ell. \quad (9)$$

Integrating (9) with respect to  $z_p$ , we obtain

$$(\Phi(c) - \Phi(-d)) \times \left[ \sum_{j=1}^{\binom{p-1}{k-1}} \int_{-d}^c \cdots \int_{-d}^c \prod_{m \in I_j^c} \Phi \left( \min_{\ell \in I_j - \{p\}} \{z_\ell + \theta_\ell\} - \theta_m \right) \prod_{\ell \in I_j - \{p\}} \phi(z_\ell) dz_\ell \right]$$

where the quantity in square [ ] brackets is exactly the coverage probability for selecting  $k-1$  out of  $p-1$  populations.

Repeating the argument, but now letting  $\theta_{p-1} \uparrow \infty$ , we obtain

$$(\Phi(c) - \Phi(-d))^2 \times \left[ \sum_{j=1}^{\binom{p-2}{k-2}} \int_{-d}^c \cdots \int_{-d}^c \prod_{m \in I_j^c} \Phi \left( \min_{\ell \in I_j - \{p, p-1\}} \{z_\ell + \theta_\ell\} - \theta_m \right) \prod_{\ell \in I_j - \{p, p-1\}} \phi(z_\ell) dz_\ell \right].$$

Continuing this process ( $k-1$  times) until the sets  $I_j$  consist of only one element, we find that the coverage probability converges to

$$(\Phi(c) - \Phi(-d))^{k-1} \left[ \sum_{j=1}^{p-k+1} \int_{-d}^c \prod_{m \in I_j^c} \Phi(z + \theta_j - \theta_m) \phi(z) dz \right], \quad (10)$$

as  $\theta_{p-k}$  increases to infinity.

Finally, we observe that the expression in square [ ] brackets in (10) corresponds to the coverage probability for selecting 1 out of  $p-k+1$  populations, and we showed in Theorem 1 that the sum is minimized at  $\theta_1 = \dots = \theta_{p-k+1}$  whenever  $d > d_0$ . We obtain the following general result that we prove in Appendix A.2.

**Theorem 2.** *Let  $X_1, \dots, X_p$  be independent random variables with  $X_i \sim N(\theta_i, \sigma^2)$  for  $i = 1, \dots, p$ . Then, for any  $c > 0$  and  $1 \leq k \leq p$  there exists  $0 < d_0 \leq c$  such that if  $d > d_0$ ,*

$$\begin{aligned} & \inf_{\theta_1, \dots, \theta_p} P(\theta_{(1)} \in (X_{(1)} - c, X_{(1)} + d), \dots, \theta_{(k)} \in (X_{(k)} - c, X_{(k)} + d)) \quad (11) \\ & = (\Phi(c) - \Phi(-d))^{k-1} [\Phi^{p-k+1}(c) - \Phi^{p-k+1}(-d)]. \end{aligned}$$

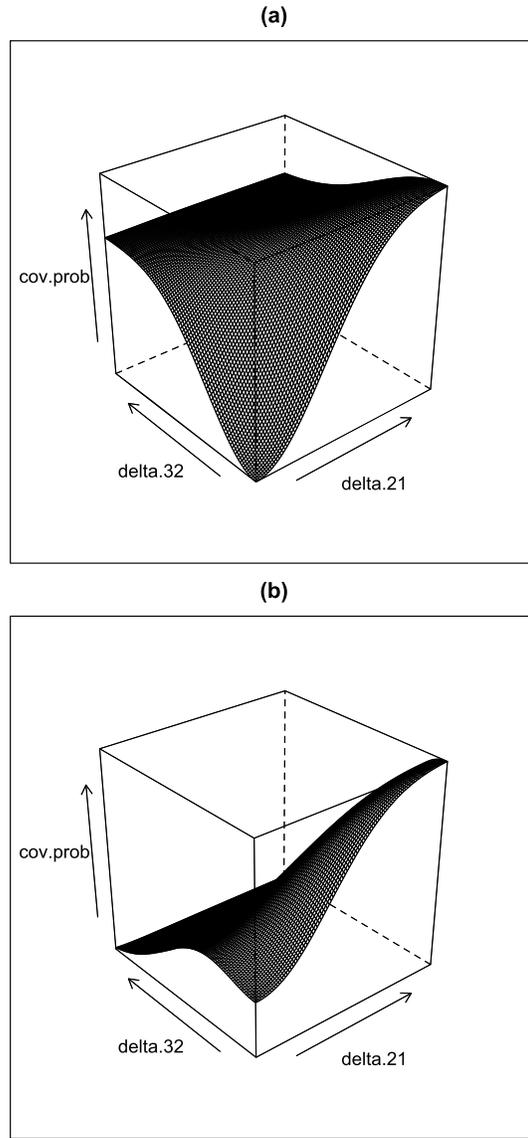


FIG 2. Coverage probability surface for the case  $p = 3$  and  $k = 1$  for  $\Delta_{32} = \theta_3 - \theta_2$  and  $\Delta_{21} = \theta_2 - \theta_1$ , with  $\theta_3 \geq \theta_2 \geq \theta_1$ . Panel (a) shows the probability surface corresponding to  $c = 1, d = 0.85$ , and panel (b) the probability surface for  $c = 1, d = 0.55$ .

### 3. Post-selection confidence intervals

For  $1 \leq i \leq p$ , let  $X_{i1}, \dots, X_{in}$  be a random sample from a population  $\pi_i$  with unknown mean  $\theta_i$  and variance  $\sigma^2$ . We assume the populations  $\pi_i$  are indepen-

dent and normally distributed or that  $n$  is large enough so we can safely apply the central limit theorem. In other words, we assume that for every  $i$ , the corresponding sample mean  $X_i = n^{-1} \sum_{j=1}^n X_{ij}$  follows a  $N(\theta_i, \sigma^2/n)$  distribution. We now apply the bounds obtained in Section 2 to obtain the desired confidence intervals. The proof of these results is a direct consequence of Theorem 2.

**Theorem 3.** *Let  $0 < \alpha < 1$  and for  $i = 1, \dots, p$ , suppose that  $X_{i1}, \dots, X_{in}$  is a random sample from a  $N(\theta_i, \sigma^2)$  population, where  $\theta_i$  is unknown, but the variance  $\sigma^2$  is known. Then, simultaneous confidence intervals for  $\theta_{(1)}, \dots, \theta_{(k)}$ , with a joint confidence coefficient of (at least)  $1 - \alpha$  are given by*

$$\left( X_{(j)} - \frac{\sigma}{\sqrt{n}}c, X_{(j)} + \frac{\sigma}{\sqrt{n}}d \right), \quad j = 1, \dots, k \leq p,$$

where  $d_0 < d \leq c$  (for some  $d_0$ ) and satisfy

$$(\Phi(c) - \Phi(-d))^{k-1} [\Phi^{p-k+1}(c) - \Phi^{p-k+1}(-d)] = 1 - \alpha. \quad (12)$$

**Corollary 1.** *Let  $0 < \alpha < 1$  and for  $i = 1, \dots, p$ , suppose that  $X_{i1}, \dots, X_{in}$  is a random sample from a  $N(\theta_i, \sigma^2)$ , where  $\theta_i$  is unknown, but  $\sigma^2$  is known. Then, a confidence interval for  $\theta_{(1)}$ , with a confidence coefficient of (at least)  $1 - \alpha$  is given by*

$$\left( X_{(1)} - \frac{\sigma}{\sqrt{n}}c, X_{(1)} + \frac{\sigma}{\sqrt{n}}d \right),$$

where  $d_0 < d \leq c$  (for some  $d_0$ ) and satisfy

$$\Phi^p(c) - \Phi^p(-d) = 1 - \alpha.$$

For given  $0 < \alpha < 1$ , the specific values of  $c, d$  can be determined numerically using the conditions in Theorem 1, optimizing for the length of the intervals. For the case  $k = 1$ , we can use equation (1) in Corollary 1, to explore the asymptotic behavior of the cutoff value  $c$ . It can be shown that the length of the intervals grow approximately at the rate  $2\sqrt{\log p}$ , as  $p$  increases. However, more interestingly, we observe that the limits of the intervals are determined by the cdf of the first order statistic of a random sample from a standard normal population and therefore, from extreme value theory results, we obtain

$$|\Phi^p(c) - \exp\{-\exp\{-(c - a_p)/b_p\}\}| \rightarrow 0, \text{ as } p \uparrow \infty,$$

where  $a_p = (2 \log p - \log \log p - \log 4\pi)^{1/2}$  and  $b_p = 1/a_p$  (see [1]). The problem of determining the asymptotic behavior of the intervals for the general case is more involved and will not be addressed in this note.

### 3.1. The unknown variance case

If the variance  $\sigma^2$  is unknown, we need to estimate its value. We assume that we have an independent estimate  $s^2$  of  $\sigma^2$ , such that  $t = s/\sigma$  has a pdf  $f(t)$ . In a

normal experiment, where we observe a sample of size  $n$  from each population,  $s^2$  can be taken as the pooled variance estimate, so that  $\nu s^2/\sigma^2 \sim \chi_\nu^2$ , a chi-square distribution with  $\nu = p(n-1)$  degrees of freedom.

For  $k = 1$  we observe that conditioning on  $t = s/\sigma$ , we can rewrite coverage probability as a mixture and obtain

$$\begin{aligned} P(\theta_{(1)} \in (X_{(1)} - sc, X_{(1)} + \lambda sc)) \\ = \int_0^\infty P(\sigma^{-1}\theta_{(1)} \in (\sigma^{-1}X_{(1)} - tc, \sigma^{-1}X_{(1)} + \lambda tc)|t)f(t)dt, \end{aligned}$$

where, for fixed  $t$ , the probability in the integrand is minimized by the expression in Corollary 1.

Using the same strategy, we can determine a lower bound for the general case and obtain the following result:

**Theorem 4.** *Let  $0 < \alpha < 1$  and for  $i = 1, \dots, p$ , suppose that  $X_{i1}, \dots, X_{in}$  is a random sample from a  $N(\theta_i, \sigma^2)$ , where  $\theta_i$  is unknown. If the variance  $\sigma^2$  is unknown, consider the estimate of  $\sigma^2$  given by  $s^2 = p^{-1} \sum_{i=1}^p s_i^2$ , with  $s_i^2 = (n-1)^{-1} \sum_{j=1}^n (X_{ij} - X_i)^2$  for  $i = 1, \dots, p$ . Then, simultaneous confidence intervals for  $\theta_{(1)}, \dots, \theta_{(k)}$ , with a joint confidence coefficient of (at least)  $1 - \alpha$  are given by*

$$\left( X_{(j)} - \frac{s}{\sqrt{n}}c, X_{(j)} + \frac{s}{\sqrt{n}}d \right), \quad j = 1, \dots, k,$$

where  $d_0 < d \leq c$  (for some  $d_0$ ) and satisfy

$$\int_0^\infty (\Phi(ct) - \Phi(-dt))^{k-1} [\Phi^{p-k+1}(ct) - \Phi^{p-k+1}(-dt)] f(t)dt = 1 - \alpha.$$

#### 4. Numerical studies

In order to evaluate the performance of the new intervals we need to look at width and coverage probability. For the case  $k = 1$ , we first look at the behavior of the actual confidence coefficient in terms of the total number of populations  $p$ . As we discussed in Section 2, the use of intervals of the form  $X_{(1)} \pm z_{\alpha/2}\sigma$  (traditional intervals) or even using a Bonferroni correction based on  $k < p$  populations will result in intervals that have a poor performance in terms of coverage probability and do not deliver the desired confidence coefficient. This feature is particularly noticeable when all the populations have similar population means, in which case, the confidence coefficient of the procedures that do not take into account the selection mechanism approaches zero rapidly.

In terms of length, in order to maintain the desired confidence coefficient the intervals need to get wider as the number of population increases. However, even for  $c = d$  (the symmetric case), the new intervals perform better than Bonferroni. Venter's method produces asymmetric intervals that are substantially shorter than Bonferroni, but optimizing the length in our approach we can obtain asymmetric confidence intervals that are consistently shorter than

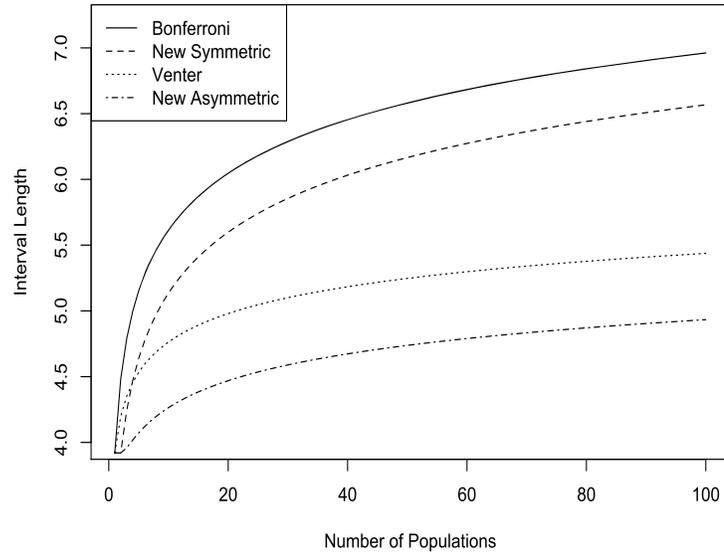


FIG 3. Length of the intervals versus number of populations for Bonferroni, Venter and the new intervals for the symmetric and asymmetric case at a nominal level of 95%.

Venter's while maintaining the desired confidence coefficient. In Figure 3 we can find a plot of the interval lengths against the number of populations for each one of these procedures for  $\alpha = 0.05$ . We observe that the new intervals are significantly shorter than the other methods, even for a small number of populations.

In applied situations, the population means  $\theta_i$  ( $1 \leq i \leq p$ ) will rarely be equal. If the largest  $\theta_i$  is far apart from the rest, then, the selected population is likely to be the population with the largest mean. When this is the case, the selection mechanism will have little or negligible effect on the outcome of the experiment and the traditional intervals will be adequate. On the other hand, by construction, the proposed intervals will have a larger confidence coefficient than the nominal level for any configuration of the  $\theta_i$  that does not correspond to the iid case. Therefore, it is important to evaluate the performance of the intervals for different configurations of the population means. We observe that the confidence coefficient of the traditional intervals is below the nominal level even when the population means are relatively separate. At the same time, the new intervals will have a coverage probability closer to the nominal level than Bonferroni and Venter intervals for any configuration. In Table 1 we summarize some of these results obtained for the case  $p = 6$ . The first column shows the true values of the population means corresponding to independent normal populations with variance one, and the remaining columns correspond to the actual coverage probabilities for the traditional intervals, Venter's, Bonferroni's and the new procedure. Notice that even when the population means are rel-

TABLE 1

Exact coverage probability of the 95% confidence intervals for the mean of the selected population obtained using the traditional approach, Venter's, Bonferroni and the new procedure for the case  $p = 6$  and different configurations of the population means.

$(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$	Trad	Venter	New	Bonf
(0,0,0,0,0,0)	0.859	0.975	0.951	0.975
(0,0.25,0.5,0.75,1,1.25)	0.870	0.976	0.953	0.976
(0,0.5,1,1.5,2,2.5)	0.896	0.979	0.960	0.979
(0,1,2,3,4,5)	0.933	0.986	0.972	0.987
(0,2,4,6,8,10)	0.949	0.983	0.967	0.991
(0,0,0,0,3,3)	0.942	0.987	0.976	0.988

atively far apart (rows 5 and 6) the traditional intervals fail to maintain the nominal coefficient and that the new intervals perform better than Bonferroni and Venter's intervals for any configuration.

For  $k > 1$  (number of selected populations) the width of the intervals will also grow as  $k$  and  $p$  increase. Of course, for fixed  $p$  the length of the new intervals will approach the length of Bonferroni's intervals when  $k$  gets closer to  $p$ , but it is important to study the behavior of the new intervals as the number of populations  $p$  increases and  $k$  is fixed or increases at a certain rate. Numerical studies show that in either case, the new intervals will be shorter than Bonferroni, for any value of  $k$ . In Figure 4 we show the results of one of these studies where we considered the number of selected population to be  $k = 5$  (fixed) regardless the total number of populations, and  $k$  (variable) equal to 5% of the total number of populations  $p$ . In both cases we observe that the proposed new intervals improve in length Bonferroni's intervals and the difference can be significant, even when the total number of populations is relatively small.

Regarding the coverage probability, we conducted numerical studies to determine the actual confidence coefficient of the intervals for different configurations of the population means. In Table 2 we summarize the results of one of these studies for the case  $p = 6$ . Again, the first column shows the true values of the population means corresponding to independent normal populations with variance one, and the remaining columns indicate the observed coverage probabilities for the corresponding number of selected populations  $k$ . The results are based on 10 replications of 25K simulations reporting and standard error  $< 0.0001$ . We observe the actual confidence coefficient is never below to the nominal level and do not exceed Bonferroni. Interestingly, the coverage probabilities get closer to the nominal level when we have some separation between the true population means (rows 5 and 6), however to establish an accurate pattern is difficult, since the actual value coverage probability depends not only on the configuration of the  $\theta_i$ , but also on  $k$  and  $p$  as can be seen in equation (8).

## 5. Discussion

A practitioner might wish to use the data to select scientifically relevant variables and then make inference on the selected ones. In this article we framework

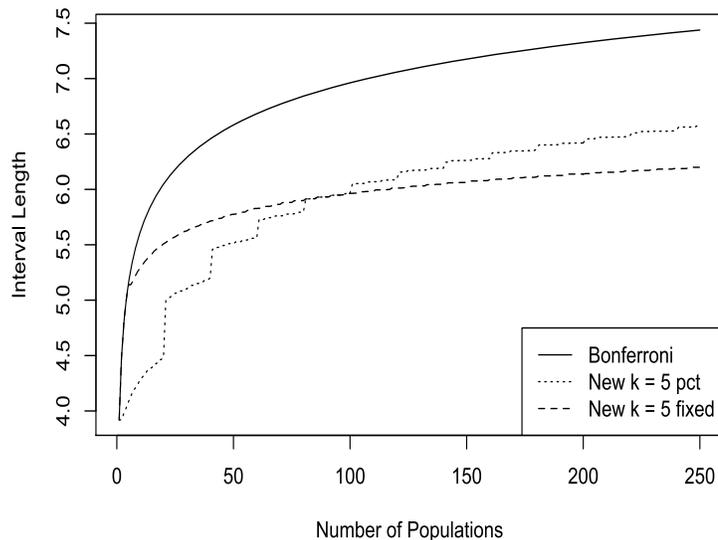


FIG 4. Length of the intervals versus number of populations for Bonferroni and the new intervals when  $k = 5$  (fixed) and  $k = 5\%$  of the total number of populations at a nominal level of 95%.

TABLE 2

Observed coverage probabilities for 95% CI for the mean of the  $k$  deleted populations using the new method and different configurations of the population means.

$(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
(0,0,0,0,0,0)	0.951	0.969	0.974	0.975	0.970	0.950
(0,0.25,0.5,0.75,1,1.25)	0.953	0.970	0.973	0.975	0.968	0.968
(0,0.5,1,1.5,2,2.5)	0.960	0.970	0.974	0.972	0.964	0.967
(0,1,2,3,4,5)	0.972	0.972	0.966	0.960	0.957	0.960
(0,2,4,6,8,10)	0.967	0.960	0.957	0.953	0.952	0.954
(0,0,0,0,3,3)	0.976	0.964	0.960	0.962	0.961	0.951

for post-selection inference by constructing simultaneous confidence intervals for the means of  $k \geq 1$  selected populations. In this context, we were able to derive a closed form exact expression for the coverage probability of such intervals and minimize it to determine a lower bound. As a result we are able to obtain simultaneous confidence intervals that do not fail to maintain the nominal confidence coefficient and retain the frequentist interpretation.

The proposed intervals do take into account the selection mechanism and therefore their length increases as the number of populations grow. Furthermore, the approach allow us to construct asymmetric intervals that correct the selection bias of the order statistics used as point estimators and shorten the length of the intervals. For  $k = 1$ , the proposed intervals perform better than Venter's and Bonferroni's intervals in both length and coverage probability.

Although the assumptions of normality and equal variances are necessary in order to obtain the exact results in closed form, the results in Theorem 2

are still valid under some mild degree of asymmetry and therefore hold for a larger family of distributions. This, in conjunction with the length of the intervals, allows the methodology to perform well under mild violations of the assumptions.

Finally, we observe that for  $k > 1$  we do not need to consider a fixed common values of  $c, d$  for all the intervals. This open the door to further investigate alternatives to optimize the length of the intervals obtained using this method and improve the results.

The approach of this article should be contrasted to the asymptotic approaches of Van de Geer & Dezeure [30] and Zhang & Zhang [34]. These articles approached the post-selection inference issue by first applying a lasso procedure to select relevant features and then use a debiased version of the lasso estimator to account the inherent randomness in the selection procedure. Symmetric intervals are then constructed based on the scaled debiased lasso estimator. The confidence intervals we construct maintain exact, not asymptotic, coverage and do not depend on the implicit selection of tuning parameters.

## Appendix A: Proof of results

### A.1. Lemma in Theorem 1

We begin by proving the following lemma which may be of independent interest:

**Lemma A.1.** *Let  $\Delta > 0$  and suppose that  $f_1(z)$  and  $f_2(z)$  are integrable functions over the real line such that  $f_1(z) > 0$  is non-decreasing and  $f_2(z) > 0$  is concave and symmetric with respect to  $\Delta/2$  and non-decreasing in  $(-\infty, \Delta/2)$ . Then, for  $g(z) = f_1(z)f_2(z)$  we have*

i) For any  $d > 0$ ,

$$\int_{\Delta-d}^{\infty} g(z)dz - \int_{-d}^{\infty} g(z)dz \leq 0$$

ii) For any  $c > 0$ ,

$$\int_{-\infty}^{\Delta+c} g(z)dz - \int_{-\infty}^c g(z)dz \geq 0$$

iii) For any  $c > 0$ , there exists  $0 < d_0 \leq c$  such that

$$a) \quad \int_{\Delta-d}^{\Delta+c} g(z)dz - \int_{-d}^c g(z)dz \geq 0, \text{ if } d_0 \leq d$$

$$b) \quad \int_{\Delta-d}^{\Delta+c} g(z)dz - \int_{-d}^c g(z)dz \leq 0, \text{ if } 0 < d \leq d_0$$

The inequalities are strict if the function  $f_1(\cdot)$  is strictly increasing in  $z$ .

*Proof.* Since  $\Delta > 0$ , cases (i) and (ii) are straightforward. For (iii), we have two possibilities for the intervals of integration:

$$\begin{aligned} -d < \Delta - d < c < \Delta + c, \text{ or} \\ -d < c < \Delta - d < \Delta + c, \end{aligned}$$

depending on whether the intervals overlap or not. Consequently, if we denote by  $R_1$  and  $R_2$  the non-common regions of integration, there are two possible cases

- 1)  $R_1 = (-d, \Delta - d)$  and  $R_2 = (c, \Delta + c)$
- 2)  $R_1 = (-d, c)$  and  $R_2 = (\Delta - d, \Delta + c)$

Regardless the case,  $R_1$  and  $R_2$  are intervals of the same length. Also, since  $f_2(z)$  is symmetric around  $\Delta/2$ , we have  $f_2(-c) = f_2(\Delta + c)$  and therefore,  $g(-c) \leq g(\Delta + c)$ . In fact, for any  $0 < \delta < \text{length}(R_1)$ , we have  $g(-c + \delta) \leq g(\Delta + c - \delta)$  and therefore

$$\int_{R_1} g(z)dz \leq \int_{R_2} g(z)dz$$

which completes the proof for part (a) when  $d_0 = 1$ . Observe that the argument is still valid for  $0 < d_0 < 1$  as long as  $f_1(z)$  is strictly increasing in some interval contained in  $R_2$ . In such case, the value of  $d_0$  will depend on the rate at which  $f_1(z)$  increases. For part (b) the inequalities reverse and the result follows completing the proof of the Lemma.  $\square$

Following a similar argument, cases (i) and (ii) in Lemma A.1 lead to the construction of one-sided intervals discussed in [32].

## A.2. Proof of Theorem 2

*Proof.* For general  $k$ , we observe that the expression for the coverage probability in (8) is non-decreasing in  $\theta_p$  for all the terms where  $\theta_p \in I_j$  and strictly decreasing in those terms where  $\theta_p \in I_j^c$ . As a result, the coverage probability will be smaller either at  $\theta_p = \theta_{p-1}$  or  $\theta_p = \infty$ . Repeating this argument for  $\theta_{p-1}, \dots, \theta_{p-k+1}$ , we obtain the coverage probability is minimized either at i)  $\theta_1 = \dots = \theta_p = \theta$  or at ii)  $\theta_1 = \dots = \theta_{p-k+1} = \theta, \theta_{p-k+2} = \dots = \theta_p = \infty$ . For each case, the corresponding coverage probabilities are:

$$\begin{aligned} i) \quad H_p(k) &= \binom{p}{k} \int_{-d}^c \dots \int_{-d}^c \Phi^{p-k}(\min\{z_1, \dots, z_k\}) \phi(z_1, \dots, z_k) dz_1 \dots dz_k \\ ii) \quad M_p(k) &= (p - k + 1) \int_{-d}^c \dots \int_{-d}^c \Phi^{p-k}(z_1) \phi(z_1, \dots, z_k) dz_1 \dots dz_k \\ &= (\Phi(c) - \Phi(-d))^{k-1} [\Phi^{p-k+1}(c) - \Phi^{p-k+1}(-d)]. \end{aligned}$$

In this context, we want to show that for any  $p > 1$ ,  $H_p(k) \geq M_p(k)$  for  $1 \leq k \leq p$ . Observe that both  $H_p(k)$  and  $M_p(k)$  are decreasing in  $k$  (they correspond to probabilities of  $k$ -simultaneous events) and that  $H_p(1) = M_p(1)$

and  $H_p(p) = M_p(p)$ . Furthermore,  $M_p(k)$  is convex as a function of  $k$  and  $H_p(k)$  is either convex, concave or changes from concave to convex. Regardless the case, it is sufficient to prove the  $H_p(p-1) \geq M_p(p-1)$  to obtain the result.

For  $j = 1, \dots, k$ , define  $R_j = \{(z_1, \dots, z_k) \in (-d, c)^k : z_j = \min\{z_1, \dots, z_k\}\}$ . Then, we can write

$$H_p(k) = \binom{p}{k} \sum_{j=1}^k \int \cdots \int_{R_j} \Phi^{p-k}(z_j) \phi(z_1, \dots, z_k) dz_1 \dots dz_k$$

$$M_p(k) = (p+1-k) \sum_{j=1}^k \int \cdots \int_{R_j} \Phi^{p-k}(z_1) \phi(z_1, \dots, z_k) dz_1 \dots dz_k$$

and taking advantage of the symmetry of the regions of integration we obtain

$$H_p(p-1) - M_p(p-1) = [p(p-1) - 2] \int \cdots \int_{R_1} \Phi(z_1) \phi(z_1, \dots, z_k) dz_1 \dots dz_{p-1}$$

$$- 2(p-2) \int \cdots \int_{R_2} \Phi(z_1) \phi(z_1, \dots, z_k) dz_1 \dots dz_{p-1}.$$

The case  $p = 2$  is straightforward. For  $p = 3$  we obtain

$$H_3(2) - M_3(2) = 4 \int \int_{R_1} \Phi(z_1) \phi(z_1) \phi(z_2) dz_1 dz_2$$

$$- 2 \int \int_{R_2} \Phi(z_1) \phi(z_1) \phi(z_2) dz_1 dz_2$$

$$= \Phi^2(c) \Phi(-d) + \Phi^3(-d) - 2\Phi^2(-d) \Phi(c)$$

$$= \Phi(-d) [\Phi^2(c) + \Phi^2(-d) - 2\Phi(c) \Phi(-d)].$$

Since  $\Phi^2(c) + \Phi^2(-d) \geq 2\Phi(c) \Phi(-d)$  for any  $c, d > 0$ , we obtain that  $H_3(2) - M_3(2) \geq 0$ . The general case follows from induction on  $p$ , completing the proof of the Theorem.  $\square$

## Acknowledgements

We would like to dedicate this article to the memory of our mentor and friend George Casella who first motivated this work and with whom we discussed many of the results here presented.

## References

- [1] AHSANULLAH, M., NEVZOROV, V. B. and SHAKIL, M. (2013). *An Introduction to Order Statistics*. Paris: Atlantis Press. [MR3025012](#)
- [2] BECHHOFER, R. E. (1954). A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. *Annals of Mathematical Statistics* **25** 16–39. [MR0060197](#)

- [3] BECHHOFFER, R. E., SANTNER, T. J. and GOLDSMAN, D. M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. New York: Wiley. [MR1341587](#)
- [4] BENJAMINI, Y. and YEKUTIELI, D. (2005). False Discovery Rate Adjusted Multiple Confidence Intervals for Selected Parameters. *Journal of the American Statistical Association* **100** 71–81. [MR2156820](#)
- [5] BERGER, J. O. (1976). Inadmissibility Results for Generalized Bayes Estimators of Coordinates of a Location Vector. *Annals of Statistics* **4** 302–333. [MR0400485](#)
- [6] BLUMENTHAL, S. and COHEN, A. (1968). Estimation of the Larger of Two Normal Means. *Journal of the American Statistical Association* **63** 861–876. [MR0231477](#)
- [7] BROWN, L. D. (1979). A Heuristic Method for Determining Admissibility of Estimators with Applications. *Annals of Statistics* **7** 960–994. [MR0536501](#)
- [8] BROWN, L. D. (1987). Personal Communication.
- [9] CHEN, H. J. and DUDEWICZ, E. J. (1976). Procedures for Fixed-Width Interval Estimation of the Largest Normal Mean. *Journal of the American Statistical Association* **71** 752–756. [MR0428573](#)
- [10] COHEN, A. and SACKROWITZ, H. B. (1982). Estimating the Mean of the Selected Population, In: *Third Purdue Symposium on Statistical Decision Theory and Related Topics*. New York: Academic Press. [MR0705292](#)
- [11] COHEN, A. and SACKROWITZ, H. B. (1986). A Decision Theoretic Formulation for Population Selection Followed by Estimating the Mean of the Selected Population, In: *Fourth Purdue Symposium on Statistical Decision Theory and Related Topics*. New York: Academic Press.
- [12] DAHIYA, R. C. (1974). Estimation of the Mean of the Selected Population. *Journal of the American Statistical Association* **69** 226–230.
- [13] EFRON, B. (2011). Tweedie’s formula and selection bias. *Journal of the American Statistical Association* **106** 1602–1614.
- [14] GUPTA, S. S. and MIESCKE, K. J. (1990). On Finding the Largest Normal Mean and Estimating the Selected Mean. *Sankhyā: The Indian Journal of Statistics, Series B* **52** 144–157.
- [15] GUPTA, S. S. and PANCHAPAKESAN, S. (2002). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Philadelphia: Society for Industrial and Applied Mathematics.
- [16] GUPTA, S. S. and SOBEL, M. (1957). On a Statistic Which Arises in Selection and Ranking Problems. *Annals of Mathematical Statistics* **28** 957–967.
- [17] GUTTMAN, I. and TIAO, G. C. (1964). A Bayesian Approach to Some Best Population Problems. *Annals of Mathematical Statistics* **35** 825–835.
- [18] HWANG, J. T. (1993). Empirical Bayes Estimation for the Means of the Selected Populations. *Sankhyā: The Indian Journal of Statistics, Series A* **55** 285–304.
- [19] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2014). Exact post-selection inference, with application to the lasso. *arXiv preprint arXiv:1311.6238*.

- [20] LELE, C. (1993). Admissibility Results in Loss Estimation. *Annals of Statistics* **21** 378–390.
- [21] PUTTER, J. and RUBINSTEIN, D. (1968). On Estimating the Mean of a Selected Population Technical Report No. 165, Department of Statistics, University of Wisconsin.
- [22] QIU, J. and HWANG, J. T. G. (2007). Sharp Simultaneous Intervals for the Means of Selected Populations with Application to Microarray Data Analysis. *Biometrics* **63** 767–776. [MR2395714](#)
- [23] REID, S. and TIBSHIRANI, R. (2014). Post selection point and interval estimation of signal sizes in Gaussian samples. *arXiv preprint arXiv:1405.3340*.
- [24] SACKROWITZ, H. and SAMUEL-CAHN, E. (1984). Estimation of the Mean of a Selected Negative Exponential Population. *Journal of the Royal Statistical Society: Series B* **46** 242–249. [MR0781883](#)
- [25] SACKROWITZ, H. and SAMUEL-CAHN, E. (1986). Evaluating the Chosen Population: A Bayes and Minimax Approach. *Lecture Notes-Monograph Series* 386–399. [MR0898261](#)
- [26] SAXENA, K. M. L. (1976). A Single-Sample Procedure for the Estimation of the Largest Mean. *Journal of the American Statistical Association* **71** 147–148. [MR0403031](#)
- [27] SAXENA, K. M. L. and TONG, Y. L. (1969). Interval Estimation of the Largest Mean of  $k$  Normal Populations with Known Variances. *Journal of the American Statistical Association* **64** 296–299.
- [28] SIMON, N. and SIMON, R. (2013). On estimating many means, selection bias, and the bootstrap. *arXiv preprint arXiv:1311.3709*.
- [29] STEIN, C. (1964). Contribution to the Discussion of Bayesian and Non-Bayesian Decision Theory. *Handout from the Institute of Mathematical Statistics Meeting*.
- [30] VAN DE GEER, B. P. R. Y. S. and DEZEURE, R. (2014). On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models. *Annals of Statistics* **42** 1166–1202. [MR3224285](#)
- [31] VENTER, J. (1988). Estimation of the Mean of the Selected Population. *Communications in Statistics-Theory and Methods* **17** 791–805. [MR0939643](#)
- [32] VENTER, J. (1988). Confidence Bounds Based on the Largest Treatment Mean. *South African Journal of Science* **84** 340–342.
- [33] VENTER, J. and STEEL, S. (1991). Estimation of the Mean of the Population Selected from  $k$  Populations. *Journal of Statistical Computation and Simulation* **38** 1–14. [MR1097121](#)
- [34] ZHANG, C. H. and ZHANG, S. (2014). Confidence Intervals for Low-Dimensional Parameters with High-Dimensional Data. *Journal of the Royal Statistical Society: Series B* **76** 217–242. [MR3153940](#)
- [35] ZHAO, Z. and HWANG, J. (2012). Empirical Bayes False Coverage Rate Controlling Confidence Intervals. *Journal of the Royal Statistical Society: Series B* **74** 871–891. [MR2988910](#)