Journal of Statistical Planning and Inference 139 (2009) 3419 – 3429

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Confidence intervals in regression utilizing prior information

Paul Kabaila*, Khageswor Giri

*Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia*

ARTICLE INFO

*Article history:*
Received 21 November 2007
Received in revised form
19 March 2009
Accepted 24 March 2009
Available online 2 April 2009

*Keywords:*
Frequentist confidence interval
Prior information
Linear regression

ABSTRACT

We consider a linear regression model with regression parameter $\beta=(\beta_1,\ldots,\beta_p)$ and independent and identically $N(0,\sigma^2)$ distributed errors. Suppose that the parameter of interest is $\theta=a^T\beta$ where $a$ is a specified vector. Define the parameter $\tau=c^T\beta-t$ where the vector $c$ and the number $t$ are specified and $a$ and $c$ are linearly independent. Also suppose that we have uncertain prior information that $\tau=0$. We present a new frequentist $1-\alpha$ confidence interval for $\theta$ that utilizes this prior information. We require this confidence interval to (a) have endpoints that are continuous functions of the data and (b) coincide with the standard $1-\alpha$ confidence interval when the data strongly contradict this prior information. This interval is optimal in the sense that it has minimum weighted average expected length where the largest weight is given to this expected length when $\tau=0$. This minimization leads to an interval that has the following desirable properties. This interval has expected length that (a) is relatively small when the prior information about $\tau$ is correct and (b) has a maximum value that is not too large. The following problem will be used to illustrate the application of this new confidence interval. Consider a $2\times 2$ factorial experiment with 20 replicates. Suppose that the parameter of interest $\theta$ is a specified *simple* effect and that we have uncertain prior information that the two-factor interaction is zero. Our aim is to find a frequentist 0.95 confidence interval for $\theta$ that utilizes this prior information.



## 1. Introduction

Consider the linear regression model $Y=X\beta+\varepsilon$, where $Y$ is a random $n$-vector of responses, $X$ is a known $n\times p$ matrix with linearly independent columns, $\beta=(\beta_1,\ldots,\beta_p)$ is an unknown parameter vector and $\varepsilon\sim N(0,\sigma^2 I_n)$ where $\sigma^2$ is an unknown positive parameter. Suppose that the parameter of interest is $\theta=a^T\beta$ where $a$ is specified $p$ vector ($a\neq 0$). Define the parameter $\tau=c^T\beta-t$ where the vector $c$ and the number $t$ are specified and $a$ and $c$ are linearly independent. Also suppose that previous experience with similar data sets and/or expert opinion and scientific background suggests that $\tau=0$. In other words, suppose that we have uncertain prior information that $\tau=0$. Of course, this includes the particular case that $c=(0,\ldots,0,1)$ and $t=0$, so that the uncertain prior information is that $\beta_p=0$. Our aim is to find a frequentist $1-\alpha$ confidence interval (i.e. a confidence interval whose coverage probability has infimum $1-\alpha$) for $\theta$ that utilizes this prior information, based on an observation of $Y$.

An attempt to incorporate the uncertain prior information that $\tau=0$ into the construction of a $1-\alpha$ confidence interval for $\theta$ is as follows. We carry out a preliminary test of the null hypothesis that $\tau=0$ against the alternative hypothesis that $\tau\neq 0$. If this null hypothesis is accepted then the confidence interval is constructed assuming that it was known *a priori* that $\tau=0$; otherwise the standard $1-\alpha$ confidence interval for $\theta$ is used. We call this the naive $1-\alpha$ confidence interval for $\theta$. This confidence interval is based on a false assumption and so we expect that its minimum coverage probability will not necessarily be $1-\alpha$. This minimum

* Corresponding author. Tel.: +61 3 9479 2594; fax: +61 3 9479 2466.
*E-mail address:* P.Kabaila@latrobe.edu.au (P. Kabaila).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.03.018

coverage probability has been investigated by Giri and Kabaila (2008), Kabaila (1998, 2005a, 2009), Kabaila and Giri (2009a) and Kabaila and Leeb (2006). In many cases this minimum is far below $1-\alpha$, showing that this confidence interval is completely inadequate. So, the naive $1-\alpha$ confidence interval fails to utilize the prior information that $\tau=0$.

Whilst the naive $1-\alpha$ confidence interval for $\theta$ fails abysmally to utilize the prior information that $\tau=0$, its form (as described in Section 2) will be used to provide some motivation for the new confidence interval described in Section 3. Similarly to Hodges and Lehmann (1952), Bickel (1983, 1984), Kabaila (1998, 2005b), Farchione and Kabaila (2008), Kabaila and Tuck (2008) and Kabaila and Giri (2009b), our aim is to utilize the uncertain prior information in the frequentist inference of interest, whilst providing a safeguard in case this prior information happens to be incorrect. We assess a $1-\alpha$ confidence interval for $\theta$ using the ratio (expected length of this confidence interval)/(expected length of standard $1-\alpha$ confidence interval). We call this ratio the scaled expected length of this confidence interval. In Section 3 we describe a new $1-\alpha$ confidence interval for $\theta$ that utilizes the prior information. This interval has endpoints that are continuous functions of the data and it has the following properties. It coincides with the standard $1-\alpha$ confidence interval when the data strongly contradict the prior information. This interval is optimal in the sense that it has minimum weighted average expected length where the largest weight is given to this expected length when $\tau=0$. This minimization leads to an interval that has the following desirable properties. This interval has scaled expected length that (a) is smaller than 1 when the prior information about $\tau$ is correct and (b) has a maximum value that is not too much larger than 1. The idea of minimizing a weighted average expected length of a confidence interval, subject to a coverage probability inequality constraint, appears to have been first used by Pratt (1961).

In Section 4 we consider the following scenario. Suppose that a $2\times 2$ factorial experiment, with factors labeled A and B and with more than one replicate, has been conducted. Also suppose that our interest is solely in the *simple* effect of changing factor A from low to high when factor B is low. Consider, for example, the case that factor A (B) being low or high corresponds to the absence or presence of treatment A (B), respectively. Our interest may be solely in the effect of treatment A compared to no treatment (cf. Hung et al., 1995). In other words, the parameter of interest $\theta$ is the *simple* effect (expected response when factor A is high and factor B is low)−(expected response when factor A is low and factor B is low). In this case, $p=4$ and we identify $\tau$ with the two-factor interaction. Suppose that previous experience with similar data sets and/or expert opinion and scientific background suggests that the two-factor interaction is zero. In a $2\times 2$ factorial clinical trial comparing two drugs whose presumed effects are on completely different systems and/or diseases, it seems reasonable to suppose that we have uncertain prior information that the two-factor interaction is zero (Stampfer et al., 1985; Steering Committee of the Physicians' Health Study Research Group et al., 1988; Buring and Hennekens, 1990; Hung et al., 1995). For an example of the elicitation of uncertain prior information in a factorial experiment via expert opinion and scientific background in a chemical context, see Dubé et al. (1996).

An attempt to utilize the uncertain prior information that the two-factor interaction is zero is to use a naive $1-\alpha$ confidence interval for $\theta$ constructed using the following preliminary test. The preliminary test is of the null hypothesis that the two-factor interaction is zero against the alternative hypothesis that the two-factor interaction is non-zero. This confidence interval has a minimum coverage probability that is far below $1-\alpha$, showing that it is completely inadequate. As an illustration, consider the case that the number of replicates is 20, $1-\alpha=0.95$ and the preliminary hypothesis test has a level of significance 0.05. We find, using the methodology of Kabaila (1998, 2005a), or Giri and Kabaila (2008) or Kabaila and Giri (2009a), that the minimum coverage probability of this confidence interval is 0.7306. The poor coverage properties of the naive confidence interval are presaged by the poor properties of some other inferences carried out after this preliminary test, see Fabian (1991), Shaffer (1991) and Ng (1994) (cf. Neyman, 1935; Bohrer and Sheft, 1979; Traxler, 1976).

The properties of the new confidence interval, described in Section 3, are illustrated in Section 4 by a detailed analysis of the $2\times 2$ factorial experiment example with 20 replicates and $1-\alpha=0.95$. Define the parameter $\gamma=\tau/\sqrt{\mathrm{var}(\hat{\tau})}$, where $\hat{\tau}$ denotes the least squares estimator of $\tau$. As proved in Section 3, the coverage probability of the new confidence interval for $\theta$ is an even function of $\gamma$. The top panel of Fig. 3 is a plot of the coverage probability of the new 0.95 confidence interval for $\theta$ as a function of $\gamma$. This plot shows that the new 0.95 confidence interval for $\theta$ has coverage probability 0.95 throughout the parameter space. As proved in Section 3, the scaled expected length of the new confidence interval for $\theta$ is an even function of $\gamma$. The bottom panel of Fig. 3 is a plot of the square of the scaled expected length of the new 0.95 confidence interval for $\theta$ as a function of $\gamma$. When the prior information is correct (i.e. $\gamma=0$), we gain since the square of the scaled expected length is substantially smaller than 1. The maximum value of the square of the scaled expected length is not too large. The new 0.95 confidence interval for $\theta$ coincides with the standard $1-\alpha$ confidence interval when the data strongly contradict the prior information. This is reflected in Fig. 3 by the fact that the square of the scaled expected length approaches 1 as $\gamma\to\infty$.

## 2. The naive confidence interval

The naive $1-\alpha$ confidence interval for $\theta$ is constructed as follows. We carry out a preliminary test of the null hypothesis that $\tau=0$ against the alternative hypothesis that $\tau\neq 0$. If this null hypothesis is accepted then the confidence interval is constructed assuming that it was known *a priori* that $\tau=0$; otherwise the standard $1-\alpha$ confidence interval for $\theta$ is used. As noted in the introduction, this confidence interval will often have minimum coverage probability far below $1-\alpha$, showing that it is completely inadequate. In this section we describe the naive confidence interval in a new form that will be used to provide some motivation for the new confidence interval described in the next section.

Let $\hat{\beta}$ denote the least squares estimator of $\beta$. Let $\hat{\Theta}$ denote $a^T\hat{\beta}$, i.e. the least squares estimator of $\theta$. Also, let $\hat{\tau}$ denote $c^T\hat{\beta}-t$, i.e. the least squares estimator of $\tau$. Define the matrix $V$ to be the covariance matrix of $(\hat{\Theta},\hat{\tau})$ divided by $\sigma^2$. Let $v_{ij}$ denote the