Journal of the KSQC
Vol. 20, No. 2, Dec. 1992

1992년 12월 韓國品質管理學會誌 제20권 제2호

Use of the Linear Placement Statistics for the Parallelism of Two Regression Lines+

Moon Sup Song*

Hong-Nai Park*

Kay O Lee**

Byeong Uk Park*

Bong Sup Shin*

Abstract

In this paper we propose a class of tests for the parallelism of two regression lines based on linear placement statistics. The results of a small-sample Monte Carlo study show that the proposed tests are reasonably good in level control and powers.

1. Introduction

Consider the regression model

$$Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij}, \quad i=1,2;\ j=1,2,\cdots N_i \tag{1.1}$$

+ This work was supported in part by the Basic Science Research Institute Program, Ministry of Eduction, 1992, Project No. BSRI-92-108

* Department of Computer Science and Statistics, Seoul, Korea

** Air Force Academy, Korea

where the x's are known constants, the α's are nuissance parameters, and the β's are regression parameters. The Y's are observable while the e's are mutually independent unobservable random variables. For line i, the e's are identically distributed according to the continuous distribution

$$P(e_{ij} \leq t) = F_i(t), \ i=1,2. \tag{1.2}$$

We wish to test the null hypothesis that two regression lines are parallel, i.e., we want to test

$$H_0 : \beta_1 = \beta_2 \tag{1.3}$$

When the error terms e_1's and e_2's are normally distributed with the same variance, the standard test derived from least-squares theory may be applied. However, some nonparametric tests which do not require any assumptions other than independent samples from identically distributed continuous populations have been considered by Hollander(1970), Pottohoff(1974) and Song(1978), among others.

Hollander(1970) proposed a distribution-free signed-rank test for testing H_0. From the viewpoint of Pitman efficiency, his test performs well for equally spaced or nearly equally spaced designs. But his method is applicable only for the case of equal numbers of observations from the two lines.

Potthoff(1974) proposed a conservative nonparametric test. His procedure is not restricted to designs with equal numbers of observations on each line. But since an upper bound of the variance is used, its performance is poor.

Trying to avoid the disadvantages of the Hollander's and Potthoff's test, Song(1978) considered a method to utilize the ordinary Wilcoxon two-sample statistic applied to the slope estimators which was introduced by Hollander(1970). Although he suggested the usage of the ordinary Wilcoxon two-sample statistic even when $F_1 \neq F_2$, but according to the Monte Carlo values in Song(1978) the probability of Type I error of the test exceeded the significance level when $F_1 \neq F_2$.

In this paper we propose a test based on the linear placement statistics which was introduced by Orban and Wolfe(1982) for two-sample problem. The class of linear placement statistics was applied to multiple comparison problems by Kin(1991). We present some results of a Monte Carlo study which compare the performance of our procedures with the above mentioned procedures.

2. The test procedures

To apply the Hollander(1970)'s pairing technique we assume that $N_1=2m$ and $N_2=2n$, discarding an observation from each sample if necessary. For line 1 pair the x_1's to form m groups of the form(x_1, x'_1), and for line 2 pair the x_2's to form n groups of the form(x_2, x'_2), each containing two unequal x's. For each group compute a slope estimator of β_i of the form$(Y-Y')/(x-x')$, where Y, Y' are observations corresponding to the paired x, x' points. The grouping cannot depend on the observed Y's but only on the x_{ij} configurations. Since the Y's for each estimator are different, $m(n)$ slope estimators are mutually independent for line 1(2). The slope estimators for different lines are also mutually independent. Thus the slope estimators of the form

$$b_{ij}=(Y_{ij}-Y_{ij'})/(x_{ij}-{}_{ij'}),\ i=1,2;j=1,\cdots,m(n) \tag{2.1}$$

are mutually independent for all $i=1,2$ and $j'<j=1,\cdots,m(n)$.

In order to construct a test statistic for the parallelism of two regression lines, we consider the linear placement statistic which was considered by Orban and Wolfe(1982). Let $U_1, U_2, \cdots U_n$ be the random variables defined by

$$mU_i=[\text{number of } b_1\text{'s} \leq b_{2i}],\ i=1,2,\cdots,n. \tag{2.2}$$

mU_i is called the placement of b_{2i} among the b_1's. Then the linear placement statistic is of the form

$$S=\sum_{i=1}^{n} \phi m(Ui) \tag{2.3}$$

where ϕ_m is the real-valued Lebesgue measurable function defined on [0,1].

The linear placement statistic that corresponds to the normal scores linear rank statistic is given by

$$S^N=\sum_{i=1}^{n} \Phi^{-1}\left(\frac{mU_i+1}{m+2}\right) \tag{2.4}$$

where Φ denotes the cdf of standard normal distribution, while an exponential scores linear placement statistic has the form

$$S^E=-\sum_{i=1}^{n} \ln\left(1-\frac{mU_i}{m+1}\right) \tag{2.5}$$

Note that if the uniform score function $\phi_m(x)=x$ is used in (2.3) then the linear placement statistic

$$S^U=\sum_{i=1}^{n} U_i \tag{2.6}$$

is equivalent to the Mann-Whitney-Wilcoxon statistic.

The one-sided test of H_0 based on (2.3) ((2.4)-(2.6)) against alternatives $\beta_2>\beta_1$ rejects H_0 if

$$S\geq s_\alpha(S^N\geq s_\alpha^*,\ S^E\geq s_\alpha^{**},\ S^U\geq s_\alpha^{\prime}) \tag{2.7}$$

where $s_\alpha(s_\alpha^*,\ s_\alpha^{**},\ s_\alpha^{\prime})$ is the upper 100αth quantile for the null distribution of $S(S^N, S^E, S^U)$.

Since the exact null distributions of the linear placements statistic are not easily obtained by recursion formuals or other direct methods, Orban and Wolfe(1982) established one sample limiting($m\rightarrow\infty$) distibution of a linear placement statistic, and derived an approximation to the exact null distribution of the statistic. They derived following facts

$$\lim_{m\rightarrow\infty} P_0(S^N\leq x)=\Phi(\frac{x}{\sqrt{n}}) \tag{2.8}$$

and

$$\lim_{m\rightarrow\infty} P_0(S^E\leq x)=G(x;n,\ 1) \tag{2.9}$$

where $G(\cdot\ ;n,\ 1)$ is the Gamma distribution function with parameters n and 1. By using (2.8) and (2.9) we can find that

$$s_\alpha^*\approx\Phi^{-1}(1-\alpha)\sqrt{\frac{n(m+n+1)}{(m+1)(m+2)}[\sum_{i=1}^{m+1}\Phi^{-2}(\frac{i}{m+2})^2]} \tag{2.10}$$

and

$$s_\alpha^{**}\approx\frac{G^{-1}(1-\alpha;n,1)=B_{n,m}}{A_{n,m}} \tag{2.11}$$

were

$$A_{n,m} = \sqrt{\frac{(m+1)(m+2)}{(m+n+1)[q_{m+1} - (m+1)\bar{\phi}^2_m}}, \quad B_{n,m} = n(1 - A_{n,m}\bar{\phi}_m$$

with

$$\bar{\phi}_m = \ell n(m+1) - \frac{\ln[(m+1)!]}{m+1}, \quad q_{m+1} = \sum_{i=1}^{m+1} [\ \ln(\frac{i}{m+1})]^2$$

The adequacies of approximation in (2.10) and (2.11) were demonstrated in Section 3 of Orban and Wolfe(1982)

3. Monte Carlo Study

In this section we compare the empirical powers and significance levels of the tests discussed in this paper by a Monte Carlo study. The proposed linear placement statistics S^N, $S^E$and S^U are compared with the parametric t-test (t), Hollander's test (H) and Potthoff's test(P).

For the model(1.1) the number of observations $N = N_1 = N_2 = 20$ is used. The design points x_{ij} are fixed with (1,2, …,20). That is, the design points are equally spaced.

We consider various underlying distributions such as uniform, normal, double exponential, Cauchy and contaminated normal distributions. The cdf of an ε-contaminated normal distribution is given by

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\sigma)$$

The computations in this Monte Carlo study are carried out on IBM PC/386-DX with numeric processor. We use the subroutine RNUN in IMSL to generate uniform random variates. The normal random variates with and wihout contamination are generated by using the sunbroutine RNNOR and the Cauchy random variates are also generated by using the subroutine.

Table 3.1 σ_i's for distributions

distribution	σ	distribution	σ
uniform	$1/\sqrt{12}$	normal	1
double expo.	$\sqrt{2}$	Cauchy	3.5
CN(.1, 5)	1.4	CN(.1, 9)	1.8

CN(ε, σ) : ε-contaminated normal

RNCHY. The inverse intergral transformation is applied to generate the double exponential random variates. The observed values of Y_{ij} are obtained from

$$Y_{ij}=\alpha_i+\beta_i x_{ij}+e_{ij},\quad i=1,2;\ j=1,2,\cdots,N.$$

For the simplicity of calculation we set the nuissance parameter $\alpha_1=\alpha_2=0$ and $\beta_1=1$. For the construction of one-sided alternatives β_2 is increased by using

$$\beta_2=\beta_1+k\left(\frac{\sigma_1+\sigma_2}{2}\right)/N$$

where $k=0$, 1, 2, 3 and the σ_i for each underlying distribution is tabulated in Table 3.1. The increment of values of k indicates the change of slopes from the null parameter space to the divergent alternatives.

For each sample generated according to the specific model, the values of the test statistics are calculated and compared with their respective critical values at significance level of $\alpha=0.053$. 500 replications are performed for each value of design constants and the empirical power of tests tabulated in Table 3.2 is given by the number of times devided by 500 that exceeds its critical value. The critical values of H and S^U are from exact table and those of S^N and S^E are calculated from (2.10) and (2.11), respectively.

Table 3.2 Empirical levels and powers

(α=.053, N=20, replication=500)

dist. of(F_1,F_2)	k	t	H	P	S^N	S^E	S^U
uniform	0	.066	.048	.026	.048	.058	.052
	1	.274	.214	.116	.214	.192	.212
	2	.534	.408	.278	.418	.396	.406
	3	.848	.678	.610	.720	.662	.690
normal	0	.052	.056	.008	.046	.048	.046
	1	.238	.170	.080	.170	.162	.170
	2	.552	.408	.298	.414	.400	.416
	3	.858	.708	.650	.720	.698	.702
double exponential	0	.054	.062	.020	.060	.074	.062
	1	.250	.220	.120	.220	.204	.218
	2	.578	.456	.410	.518	.488	.522
	3	.844	.712	.732	.742	.698	.752
contaminated normal (ε=.1, σ=5)	0	.040	.042	.018	.062	.058	.070
	1	.206	.202	.104	.222	.214	.222
	2	.426	.390	.370	.466	.414	.480
	3	.712	.624	.700	.716	.676	.730
contaminated normal (ε=.1, σ=9)	0	.044	.040	.018	.064	.064	.070
	1	.188	.224	.160	.252	.254	.262
	2	.366	.416	.502	.546	.506	.574
	3	.606	.624	.780	.770	.732	.792
Cauchy	0	.032	.038	.016	.044	.052	.040
	1	.186	.192	.202	.264	.234	.268
	2	.304	.400	.552	.514	.482	.554
	3	.458	.562	.816	.740	.682	.774
normal	0	.036	.048	.014	.066	.142	.054
	1	.194	.238	.218	.338	.464	.308

-Cauchy	2	.396	.442	.546	.664	.820	.622
	3	.506	.634	.820	.830	.926	.802
Cauchy -double-expo.	0	.042	.052	.020	.046	.030	.052
	1	.176	.218	.198	.240	.180	.278
	2	.380	.396	.526	.490	.336	.570
	3	.538	.612	.818	.746	.572	.778
normal -con. nor. (.1, 9)	0	.058	.070	.020	.074	.084	.070
	1	.218	.230	.148	.266	.296	.254
	2	.456	.476	.454	.564	.636	.550
	3	.694	.672	.804	.836	.876	.834
Cauchy -con. nor. (.1, 9)	0	.046	.054	.024	.046	.036	.050
	1	.172	.182	.182	.224	.152	.270
	2	.360	.428	.564	.530	.442	.590
	3	.520	.600	.812	.736	.634	.796

By inspecting Table 3.2, we can find that the empirical significance level of linear placement statistics is relatively good except S^E when $F_1 \neq F_2$ and Potthoff's P test is too conservative for potential users. For short or medium tailed distributions such as uniform or normal the empirical powers of the proposed tests are lower than the parametric t-test, but moderately higher than nonparametric competitors. For heavy tailed distributions the proposed tests perform better than the other tests in the most cases. Our proposed tests also perform well for the case that unequal distribution is used in each line.

Refferences

1. Hollander, M.(1970). "A distribution-free test for parallelism", Journal of the American Statistical Association 65, 387-394.
2. Kim, D.J.(1991). "Distribution-free tests based on placements and partially sequential treatments versus control procedures", unpublished doctoral dissertation, The Ohio State University, Columbus, Ohio.
3. Orban, J. and Wolfe, D.A.(1982). "A class of distribution-free two-sample tests based on placements", Journal of the American Statistical Association 77, 666-672.
4. Potthoff, F.(1974). "A nonparametric test of whether two simple regression lines are parallel", The Annals of Statistics 2, 295-310.
5. Song, M.S.(1978). "Use of the Wilcoxon test for the parallelism of two regression lines", Journal of the Korean Mathematical Society 14, 223-228.