DOI: 10.21276/sjpms

**Scholars Journal of Physics, Mathematics and Statistics** **ISSN 2393-8056 (Print)**
Sch. J. Phys. Math. Stat. 2017; 4(3):102-108 **ISSN 2393-8064 (Online)**

(An International Publisher for Academic and Scientific Resources)

# Empirical likelihood confidence intervals on the mean differences after inverse probability weighted imputation

**Lirong Wang, Hui Liu**
School of Information , Hunan University of Humanities, Science and Technology, Hunan Loudi, 417000,P.R.China

***Corresponding Author:**
Lirong Wang
Email: ldwlr1234@163.com

**Abstract:** Consider two linear models with missing data, where the covariates are not missing and response variables are missing at random (MAR). The inverse probability weighted imputation is used to impute the missing data of response variables. We construct the empirical log-likelihood ratios on mean differences of two response variables. And the asymptotic distributions for the empirical log-likelihood ratios of mean differences of response variables are standard $\chi_1^2$ comparing with the results of previous studies. The empirical likelihood confidence intervals for mean differences of response variables is more accurate because the errors caused right of the coefficient estimates is reduced.
**Keywords:** missing data; inverse probability weighted imputation; mean; empirical likelihood; confidence intervals
**AMS 2000 subject classifications:** Primary 62G05; secondary 62E20

## INTRODUCTION

In data mining and machine learning, identifying the mean differences between two populations is useful in predicting the properties of a group using one another [1]. For example, in medical research, it is an important means to compare the mean value of prolonging patient's life between a group using a new product (medicine) and a group with another product. In this paper we study empirical likelihood confidence intervals for mean differences between two linear models with missing data.

Consider the following two linear regression models:

$$X = U'\beta + W_x(X)\varepsilon, \quad Y = V'\rho + W_y(Y)\tau,$$

where $X, Y$ are scalar response variables, $U, V$ are $p \times 1$ and $q \times 1$ vector of random design variables, regression parameters $\beta \in R^p, \rho \in R^q$, $W_x(X)$ and $W_y(Y)$ are strictly positive known functions, errors $\varepsilon$ and $\tau$ are random errors with $E(\varepsilon \mid X) = E(\tau \mid Y) = 0$ , $0 < Var\,\varepsilon_i = \sigma_\varepsilon^2 < \infty, 0 < Var\,\tau_j = \sigma_\tau^2 < \infty$. Suppose that we have incomplete i.i.d. observations from the two models:

$$Z_{x_i} = (x_i, u_i, \delta_{x_i}), i = 1, \cdots, m; \quad Z_{y_j} = (y_j, v_j, \delta_{y_j}), j = 1, \cdots, n,$$

where all $\{u_i, i = 1, \cdots, m\}$, $\{v_j, j = 1, \cdots, n\}$ are observed, $\{x_i, i = 1, \cdots, m\}$, $\{y_j, j = 1, \cdots, n\}$ are incomplete, and $\delta_{x_i} = 0$ if $x_i$ is missing, $\delta_{y_j} = 0$ if $y_j$ is missing, otherwise. Throughout this paper, we assume that $X, Y$ are missing at random (MAR). That is,

$$P(\delta_x = 1 \mid x, u) = P(\delta_x = 1 \mid u) = \pi(u), P(\delta_y = 1 \mid y, v) = P(\delta_y = 1 \mid v) = \pi^*(v).$$

Denote the distribution function of $X$ is $F$ , the distribution function of Y is $G$ , mean of X is $\theta(\theta = EX)$, mean of Y is $EY$ , $\Delta = EY - EX$ , that is $EY = \Delta + \theta$ .

In this paper, we focus on constructing EL confidence regions on mean differences ($\Delta$) of response variables $X$ and $Y$ in two linear models with missing data. The EL ratio statistics are constructed based on the inverse probability weighted imputation approach, which asymptotically have $\chi_1^2 - tape$ distributions.

Our work is closely related with Wang and Rao [2-4], Xue [5], Qin and Lei [6], but it is different in a number of ways, and the research contents have not been studied at present. Wang and Rao [3] developed an imputed empirical likelihood (EL) method to construct confidence intervals for the mean. The main idea is to impute the missing values by their predicted values. Then a complete data EL method is used from the imputed data set as if they were i.i.d. observations. But the EL ratio statistic for $EY$ has a limiting distribution of a scaled $\chi^2$ with unknown weight, which cannot be applied directly to make inference for mean. An adjusted EL is thus needed to obtain a confidence interval for mean. This also would lead to a loss of the accuracy of the confidence interval. To solve this problem, Xue [5] combined the EL method and the inverse probability weighted imputation technique to study the construction of confidence intervals and regions for mean. It is shown that the EL ratios based on the inverse probability weighted imputation are asymptotically standard chi-squared, which can be used directly to construct confidence intervals and regions for mean. This is a nice feature. However, somewhat strong conditions are required, which restrict the applicable scope of the approach. Then Qin and Lei [6] constructed empirical likelihood statistics on mean which have the $\chi_1^2 - tape$ limiting distributions under some new conditions based on Xue [5], their results broaden the applicable scope of the approach combined with Xue [5].

The inverse weighted approaches are widely used in the situation of missing covariates [7, 8]. The EL method to construct confidence intervals, proposed by Owen [9-11], has many advantages over its counterparts like the normal-approximation-based method and the bootstrap method [12, 13]. In particular, the prior constraints on region shape are not needed to impose, the construction of a pivotal quantity is not required, and the constructed region is range preserving and transformation-respecting.

The rest of the paper is organized as follows. In Section 2, we introduce two imputation methods: linear regression imputation and inverse probability weighted imputation. In Section 3, we develop EL approach based on the inverse probability weighted imputation technique, and show that the resulting empirical log-likelihood is asymptotically standard chi-squared under some conditions, which is used to obtain EL confidence intervals on mean differences. A short discussion on bandwidth selection is given in Section 4. Section 5 reports some simulation results to study the performance of the proposed confidence intervals.

**Imputation methods**

To implement regression imputation, we need an initial estimators of $\beta$ and $\rho$ first. Then two imputation methods are introduced, namely, linear regression imputation and inverse probability weighted imputation.

**Initial estimator of $\beta$ and $\rho$**

Based on the completely observed pairs $(x_i, u_i),\ i = 1, \cdots, m;\ (y_j, v_j),\ j = 1, \cdots, n$, define the weighted least square estimators of $\beta$ and $\rho$ respectively as follows

$$\hat{\beta}_r = \left(\sum_{i=1}^{m} \frac{\delta_{x_i} u_i u_i'}{W_x^2(x_i)}\right)^{-1} \sum_{i=1}^{m} \frac{\delta_{x_i} u_i x_i}{W_x^2(x_i)}, \quad \hat{\rho}_r = \left(\sum_{j=1}^{n} \frac{\delta_{y_j} v_j v_j'}{W_y^2(y_j)}\right)^{-1} \sum_{j=1}^{n} \frac{\delta_{y_j} v_j y_j}{W_y^2(y_j)}.$$

**Linear regression imputation**

For missing $x_i$, $y_j$ respectively use the predicted response of $u_i$ and $v_j$ to impute is a commonly used method, i.e. use $x_i^*$ and $y_j^*$

$$x_i^* = u_i'\hat{\beta}_r, \quad y_j^* = v_j'\hat{\rho}_r$$

to impute missing $x_i$ and $y_j$. Then two 'complete' data sets for $x_i$, $y_j$ are obtained as $\{x_I, i = 1, \cdots, m\}$ and $\{y_I, j = 1, \cdots, n\}$, i.e.

$$x_I = \delta_{x_i} x_i + (1 - \delta_{x_i}) x_i^*, \quad y_I = \delta_{y_j} y_j + (1 - \delta_{y_j}) y_j^*.$$

**Inverse probability weighted imputation**

Denote the probability that $X$ is not missing by $\pi(u)$ given $u$, and $Y$ is not missing by $\pi^*(v)$ given $v$, i.e. $\pi(u) = P(\delta_x = 1 | u), \pi^*(v) = P(\delta_y = 1 | v)$, We also denote $\pi(u_i) = \pi_i, \pi^*(v_j) = \pi^*_j$. We use $(\delta_{x_i}/\pi_i)x_i + (1-\delta_{x_i}/\pi_i)u_i'\hat{\beta}_r, i = 1,\cdots,m$ as 'complete' data set for $X$ if all $\pi_i$ are known. Similar to the impute of $X$ we use $(\delta_{y_j}/\pi^*_j)y_j + (1-\delta_{y_j}/\pi^*_j)v_j'\hat{\rho}_r, j = 1,\cdots,n$ as 'complete' data set for $Y$ if all $\pi^*_j$ are known, which can be viewed as the combination of Horvitz-Thompson inverse-selection weighted method and imputation method. But in general, $\pi_i's$ and $\pi^*_j's$ are unknown, in this case we usually adopt the weight function method to estimate them. Take $0 < h = h_m \to 0, 0 < h^* = h^*_n \to 0$, and two nonnegative kernel functions $k(u), k^*(v), u \in R^p, v \in R^q$. Let $k_h(u) = k(u/h), k^*_{h^*}(v) = k^*(v/h^*)$. Then $\pi_i's$ and $\pi^*_j's$ are estimated respectively by $\hat{\pi}(u) = \sum_{i=1}^m W_{mi}(U)\delta_{x_i}$ and $\hat{\pi}^*(v) = \sum_{j=1}^n \tilde{W}_{nj}(V)\delta_{y_j}$. where $W_{mi}(U)$ and $\tilde{W}_{nj}(V)$ are the Nadaraya-Watson weight with

$$W_{mi}(U) = k_h(u-u_i)/\sum_{l=1}^m (u-u_l), \tilde{W}_{nj}(V) = k^*_{h^*}(v-v_j)\sum_{l^*=1}^n k^*_{h^*}(v-v_{l^*}).$$

So we can obtain the 'complete' sets for $X$ and $Y$ as follows

$$x_{I,i} = \frac{\delta_{x_i}}{\hat{\pi}_i}x_i + (1-\frac{\delta_{x_i}}{\hat{\pi}_i})u_{i'}\hat{\beta}_r, y_{I,j} = \frac{\delta_{y_j}}{\hat{\pi}^*_j}y_j + (1-\frac{\delta_{y_j}}{\hat{\pi}^*_j})v_{j'}\hat{\rho}_r.$$

Wang and Rao [3] developed an imputed EL method to construct confidence intervals for the mean $EX$, where the 'complete' set for $X$ was obtained under linear regression imputation, and the EL ratio statistic for $EX$ has a limiting distribution of a scaled $\chi^2_1$ with unknown weight. Qin and Lei [6] constructed confidence intervals on the mean $EX$ using inverse probability weighted imputation, and a scaled $\chi^2_1 - tape$ limiting distribution of the EL ratio statistic for $EX$ is obtained. In this paper, we will use the inverse probability weighted imputation to impute the missing data of $X$ and $Y$, and construct the EL statistics on mean differences of $X$ and $Y$, then show that the EL statistics have a $\chi^2_1 - tape$ limiting distributions, which are used to construct EL confidence intervals without adjustment.

**The EL confidence intervals**

Throughout this paper, we define 0/0 as 0. Let

$$\omega_1(x_{I,i},\theta,\Delta) = x_{I,i} - \theta, \quad \omega_2(y_{I,j},\theta,\Delta) = y_{I,j} - \theta - \Delta,$$

Similar to Owen (1990), we define the empirical log-likelihood ratio on $\Delta$ as follows

$$\prod_{i=1}^m p_i \prod_{j=1}^n q_j,$$

where $p_i > 0, i = 1,\cdots,m, \sum_{i=1}^m p_i = 1$, $q_j > 0, j = 1,\cdots n, \sum_{j=1}^n q_j = 1$, and such that $p_i > 0, i = 1,\cdots,m, \sum_{i=1}^m p_i\omega_1(x_{I,i},\theta,\Delta) = 0$, $q_j > 0, j = 1,\cdots n, \sum_{j=1}^n q_j\omega_2(y_{I,j},\theta,\Delta) = 0$. So there are the EL statistics as

$$R(\Delta) = \sup_{p_i>0, i=1,\cdots,m, q_j>0, j=1,\cdots,n,\theta} \{\sum_{i=1}^m \log(mp_i) + \sum_{j=1}^n \log(nq_j)\} = \sup_\theta R(\Delta,\theta),$$

where

$$R(\Delta,\theta) = \sup_{p_i>0,\, i=1,\cdots m,\, q_j>0,\, j=1,\cdots,n,} \{\sum_{i=1}^{m} \log (mp_i) + \sum_{j=1}^{n} \log (nq_j)\}$$

It can be shown, by using the Lagrange multiplier method, that

$$R(\Delta,\theta) = -\sum_{i=1}^{m} \log \{1+\lambda_1(\theta)\omega_1(x_{I,i},\theta,\Delta)\} - \sum_{j=1}^{n} \log \{1+\lambda_2(\theta)\omega_2(y_{I,j},\theta,\Delta)\}$$

where $\lambda_j(\theta),\ j = 1,2,$ are the solutions of the equations

$$\frac{1}{m}\sum_{i=1}^{m} \frac{\omega_1(x_{I,i},\theta,\Delta)}{1+\lambda_1(\theta)\omega_1(x_{I,i},\theta,\Delta)} = 0$$

and

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\omega_2(y_{I,j},\theta,\Delta)}{1+\lambda_2(\theta)\omega_2(y_{I,j},\theta,\Delta)} = 0.$$

Let $\partial R(\theta,\Delta)/\partial\theta = 0$, then we can obtain the EL equation

$$\frac{1}{m}\sum_{i=1}^{m} \frac{\alpha_1(x_{I,i},\theta,\Delta)}{1+\lambda_1(\theta)\omega_1(x_{I,i},\theta,\Delta)} + \frac{1}{n}\sum_{j=1}^{n} \frac{\alpha_2(y_{I,j},\theta,\Delta)}{1+\lambda_2(\theta)\omega_2(y_{I,j},\theta,\Delta)} = 0 \qquad (1)$$

where

$$\alpha_1(x_{I,i},\theta,\Delta) = \frac{\delta_{x_i}}{\pi(u_i)}\frac{\partial\omega_1(x_{I,i},\theta,\Delta)}{\partial\theta} + (1-\frac{\delta_{x_i}}{\pi(u_i)})\frac{\partial\omega_1(x_{I,i},\theta,\Delta)}{\partial\theta} \quad i = 1,\cdots,m,$$

$$\alpha_2(y_{I,j},\theta,\Delta) = \frac{\delta_{y_j}}{\pi^*(v_j)}\frac{\partial\omega_2(y_j,\theta,\Delta)}{\partial\theta} + (1-\frac{\delta_{y_j}}{\pi^*(v_j)})\frac{\partial\omega_2(y_j^*,\theta,\Delta)}{\partial\theta}, \quad j = 1,\cdots,n.$$

Use $\|u\|$ and $\|v\|$ to denote the $L_2-norm$ in $R^p$ and $R^q$ respectively. Assume that the probability density functions of $U$ and $V$ exist, and $u: f(\cdot)$ and $v: g(\cdot)$. Use $\theta_0$ to denote the true value of $\theta$, and $\theta_0 \in \Omega$ and $\Omega$ is an open interval. Similar to Qin and Lei [6] we give some regularity conditions needed of the results in this paper.

(C1) $f(u)$ and $g(v)$ all are bounded, and there exist constants $a, b > 0,\ a', b' > 0$ such that

$$\int_{x\in S(u,r)\bigcap A} f(x)dx \ge ar^p, \quad \int_{y\in S(v,r')\bigcap A'} f(y)dy \ge a'r'^q,$$

for all $r\in[0,b], r'\in[0,b']$ and all $u\in A, v\in A'$, where $A$ and $A'$ are the supports of $U$ and $V$ respectively, $S(u,r)$ is the closed sphere (under the $L_2-norm$ ) with center $u$ and radius $r$, $S(v,r')$ is the closed sphere (under the $L_2-norm$ ) with center $v$ and radius $r'$.

(C2)The probability function $\pi(u)$ is uniformly continuous on $A$, $\pi^*(v)$ is uniformly continuous on $A'$, and there exist some positive constant $C_0>0,\ C_0'>0$ such that $min_{1\le i\le m}\pi(u_i)\ge C_0$, and $min_{1\le j\le n}\pi^*(v_j)\ge C_0'$ a.s.

(C3) $K$ and $K^*$ are bounded probability density functions with bounded supports, and all satisfie the Lipschitz condition. There exist positive constants $C_1, C_2, C_1', C_2', \eta, \eta'$, such that,

$$C_1 I(\|u_0\|\le\eta) \le K(u_0) \le C_2 I(\|u_0\|\le\eta),\ C_1' I'(\|v_0\|\le\eta') \le K^*(v_0) \le C_2' I'(\|v_0\|\le\eta'),$$

where $I$, $I'$ are the indicator functions.

(C4) $h\to 0,\ mh^p/\log m\to\infty$ as $m\to\infty$, $h'\to 0,\ nh'^q/\log n\to\infty$ as $n\to\infty$.

(C5) $E(\varepsilon \mid u) = 0,\ E(|\varepsilon|)^3 < \infty, E(\|u\|^3) < \infty, \sigma_0^2 > 0, \sum_0 > 0$,
$E(\tau \mid v) = 0,\ E(|\tau|)^3 < \infty, E(\|v\|^3) < \infty, \sigma_0'^2 > 0, \sum'_0 > 0$,
where

$$\sigma_0^2 = E\{\sigma^2(u)W_x^2(u)/\pi(u)\} + Var\{m(u)\},\ \sum_0 = E\{\sigma^2(u)W_x^2(u)uu'/\pi(u)\},$$

$$\sigma_0'^2 = E\{\sigma'^2(v)W_y^2(v)/\pi^*(v)\} + Var\{m(v)\},\ \sum'_0 = E\{\sigma'^2(v)W_y^2(v)vv'/\pi^*(v)\},$$
$$m(u) = u'\beta, m(v) = v'\rho, \sigma^2(u) = E(\varepsilon^2 \mid u), \sigma'^2(v) = E(\tau^2 \mid v).$$

**Theorem:** *Suppose that assumptions (C1) through (C5) are satisfied. Then there exists a root* $\theta_{m,n}$ *of equation (3.1) such that* $R(\Delta,\theta)$ *attains its local maximum at* $\theta_{m,n}$*, and, as* $m, n \to \infty$ ,

$$\sqrt{m}(\theta_{m,n} - \theta_0) \xrightarrow{d} N(0, \frac{\sigma_0^2\sigma_0'^4\beta_x^2 + k\sigma_0^4\sigma_0'^2\beta_y^2}{c_0^2}),$$
$$-2R(\Delta,\theta_{m,n}) \xrightarrow{d} \chi_1^2,$$

where $\sigma_0, \sigma_0'$ are defined in (C5), $k = m/n$,

$$\beta_x = E\alpha_1(x_{I,i},\theta_0,\Delta), \beta_y = E\alpha_2(y_{I,j},\theta_0,\Delta), c_0 = k\sigma_0^2\beta_y^2 + \sigma_0'^2\beta_x^2,$$

$\chi_1^2$ is a chi-squared distribution with one degree of freedom.

Let $t_\alpha$ satisfy $P(\chi_1^2 \le t_\alpha) = 1 - \alpha$ . It follows from Theorem 3.1 that an EL based confidence interval on $\Delta$ with asymptotically correct coverage probability $1 - \alpha$ can be constructed as

$$\{\Delta : R(\Delta,\theta_{m,n}) \le t_\alpha\}.$$

**Bandwidth selection**

Cross-validation method in choosing bandwidths is recommended. We select $h$ and $h^*$ by minimizing

$$CV(h) = \sum_{i=1}^{m}\{\delta_{x_i} - \hat{\pi}_{-i}(u_i)\},\ C'V(h^*) = \sum_{j=1}^{n}\{\delta_{y_j} - \hat{\pi}^*_{-j}(v_j)\},$$

where $\hat{\pi}_{-i}(\cdot)$ and $\hat{\pi}^*_{-j}(\cdot)$ respectively are 'leave out' versions of $\hat{\pi}(\cdot)$ and $\hat{\pi}^*(\cdot)$ .

**Simulations**

We conducted a small simulation study on the finite sample performance of the EL confidence intervals of $\Delta$ proposed in Section 5. We used the models

$$x = 2u'\beta + |u|^{1/2}\ \varepsilon, \beta = 3,\ u : N(1.5,3),\quad \varepsilon : N(0,1),$$
$$y = 2v'\rho + |v|^{1/3}\ \tau, \rho = 3,\ v : N(2,5),\quad \tau : N(0,1).$$

The weight functions $W_{mi}(U)$ and $\tilde{W}_{nj}(V)$ used in Section 2 were chosen as

$$W_{mi}(U) = k_h(u - u_i)/\sum_{l=1}^{m}(u - u_l), \tilde{W}_{nj}(V) = k^*_{h^*}(v - v_j)\sum_{l^*=1}^{n}k^*_{h^*}(v - v_{l^*}).$$

where $k_h(u) = (2/3)(2 \ - u^2)^2 I(|u| \le 1),\ k^*_{h^*}(v) = (2/3)(2 \ - v^2)^2 I(|v| \le 1)$ . $h$ and $h^*$ were chosen by the cross-validation method introduced in Section 4.

We considered the following four cases of response probabilities under the MAR assumption:

$$case\ 1.\ \pi_1(u)=\begin{cases}0.75+0.25\,|u-1|, & if\ |u-1|\le 3.5,\\ 0.95, & elsewhere\end{cases},\quad \pi_1^*(v)=\begin{cases}0.9+0.1\,|v-1|, & if\ |v-1|\le 3.5,\\ 0.95, & elsewhere\end{cases}.$$

$$case\ 2.\ \pi_2(u)=\begin{cases}0.8+0.2\,|u-1|, & if\ |u-1|\le 3,\\ 0.9, & elsewhere\end{cases},\quad \pi_2^*(v)=\begin{cases}0.82+0.18\,|v-1|, & if\ |v-1|\le 3,\\ 0.9, & elsewhere\end{cases}.$$

$$case\ 3.\ \pi_3(u)=\begin{cases}0.91+0.09\,|u-1|, & if\ |u-1|\le 4.2,\\ 0.945, & elsewhere\end{cases},\quad \pi_3^*(v)=\begin{cases}0.91+0.09\,|v-1|, & if\ |v-1|\le 4.2,\\ 0.945, & elsewhere\end{cases}.$$

$$case\ 4.\ \pi_4(u)=\pi_4^*(v)=0.85.$$

For each of the four cases, we generated 3000 random samples of incomplete data

$$\{(x_i,u_i,\delta_{x_i}),(y_j,v_j,\delta_{y_j}),\ i=1,\cdots,m,\ j=1,\cdots,n\},$$

for $(m,n)=\{(100,100), (150,100), (150,150), (200,150), (200,200), (300,250)\}$ from the model and specified response probability functions. For nominal confidence level $1-\alpha=0.95$. Using the simulated samples, we evaluated the coverage probability (CP) and average length (AL) of the EL based confidence intervals on $\Delta$ proposed in Section 3. Table 1-4 repot the simulation results for $\Delta$ under different cases respectively.

| **Table-1: The simulation rexults under case 1** | | | | **Table-2: The simulation rexults under case 2** | | | |
|---|---|---|---|---|---|---|---|
| $(\pi,\pi^*)$ | (m,n) | AL | CP(%) | $(\pi,\pi^*)$ | (m,n) | AL | CP(%) |
| $(\pi_1,\pi_1^*)$ | (100,100) | 0.66091 | 93.4 | $(\pi_2,\pi_2^*)$ | (100,100) | 0.68211 | 92.8 |
| | (150,100) | 0.61369 | 93.8 | | (150,100) | 0.50028 | 94.0 |
| | (150,150) | 0.54324 | 94.1 | | (150,150) | 0.53480 | 94.3 |
| | (200,150) | 0.55990 | 95.7 | | (200,150) | 0.54661 | 95.6 |
| | (200,200) | 0.54522 | 95.1 | | (200,200) | 0.51006 | 95.5 |
| | (300,250) | 0.50012 | 95.3 | | (300,250) | 0.48002 | 96.0 |

| **Table-3: The simulation rexults under case 3** | | | | **Table-4: The simulation rexults under case 4** | | | |
|---|---|---|---|---|---|---|---|
| $(\pi,\pi^*)$ | (m,n) | AL | CP(%) | $(\pi,\pi^*)$ | (m,n) | AL | CP(%) |
| $(\pi_3,\pi_3^*)$ | (100,100) | 0.68993 | 92.8 | $(\pi_4,\pi_4^*)$ | (100,100) | 0.71133 | 93.4 |
| | (150,100) | 0.54399 | 93.7 | | (150,100) | 0.51055 | 94.6 |
| | (150,150) | 0.55002 | 94.6 | | (150,150) | 0.56511 | 94.5 |
| | (200,150) | 0.60011 | 94.9 | | (200,150) | 0.52108 | 95.8 |
| | (200,200) | 0.54667 | 95.2 | | (200,200) | 0.48992 | 95.7 |
| | (300,250) | 0.47802 | 95.1 | | (300,250) | 0.54005 | 95.0 |

The simulation results in tables 1-4 reveal the following facts:

- For every response rate and sample size, the coverage probabilities (CP) of EL confidence intervals are close to the nominal level 0.95, and the average lengths of intervals are small.
- The coverage probabilities (CP) of EL confidence intervals go closer to the nominal level 0.95 as the sample size increases.
- The coverage probabilities (CP) of EL confidence intervals go closer to the nominal level 0.95 as the response rate becomes higher.
- In almost all situations, the average lengths (AL) of EL confidence intervals become smaller as the sample size increases.
- In almost all situations, the average lengths (AL) of EL confidence intervals become smaller as the response rate

becomes higher.

**ACKNOWLEDGEMENT:**
This work was supported in part by Project of Hunan Province Education Department(NO. 14C0609).

## REFERENCES

1. Owen A. Data squashing by empirical likelihood. Data Mining and Knowledge Discovery. 2003 Jan 1;7(1):101-13.
2. Wang Q, Rao JN. Empirical likelihood for linear regression models under imputation for missing responses. Canadian Journal of Statistics. 2001 Dec 1;29(4):597-608.
3. Wang Q, Rao JN. Empirical Likelihood-based Inference in Linear Models with Missing Data. Scandinavian Journal of Statistics. 2002 Sep 1;29(3):563-76.
4. Wang Q, Rao JN. Empirical likelihood-based inference under imputation for missing response data. The Annals of Statistics. 2002;30(3):896-924.
5. Xue L. Empirical likelihood for linear models with missing responses. Journal of Multivariate Analysis. 2009 Aug 31;100(7):1353-66.
6. Qin Y, Lei Q. On empirical likelihood for linear models with missing responses. Journal of statistical planning and inference. 2010 Nov 30;140(11):3399-408.
7. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. Journal of the American statistical Association. 1994 Sep 1;89(427):846-66.
8. Liang H, Wang S, Robins JM, Carroll RJ. Estimation in partially linear models with missing covariates. Journal of the American Statistical Association. 2004 Jun 1;99(466):357-67.
9. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. Biometrika. 1988 Jun 1;75(2):237-49.
10. Owen A. Empirical likelihood ratio confidence regions. The Annals of Statistics. 1990 Mar 1:90-120.
11. Owen AB. Empirical Likelihood, NewYork: Chapman and Hall; 2001.
12. Hall P, La Scala B. Methodology and algorithms of empirical likelihood. International Statistical Review/Revue Internationale de Statistique. 1990 Aug 1:109-27.
13. Hall P. The Bootstrap and Edgeworth Expansion, NewYork: Springer; 1992.