김일성종합대학학보 JOURNAL OF KIM IL SUNG UNIVERSITY
(자연과학) (NATURAL SCIENCE)
주체105(2016)년 제62권 제11호 Vol. 62 No. 11 JUCHE105(2016).

정규모집단에 대한 미래 m 개관측의 순서통계량과 표본범위의 예측구간

안정화, 한광룡

통계적예측문제의 중요성으로부터 론문에서는 정규모집단에 대한 미래 m 개관측의 순서통계량과 표본범위의 예측구간에 대하여 론의하였다.

선행연구[2]에서는 정규분포에 대한 미래 m 개관측전체와 평균, 표준편차를 포함하는 예측구간들을 론의하였으며 선행연구[1]에서는 통계적구간으로서의 믿을구간과 예측구간, 허용구간들에 대하여 론의하였다. 또한 선행연구[5]에서는 임의의 주어진 분포에 대한 한가지 예측구간구성법을 제기하고 정규분포에 대한 미래 m 개관측의 순서통계량의 $1-\alpha$ 아래예측한계를 얻기 위하여 순서통계량의 기준예측량을 생각하고 이와 관련하여 구성되는 통계량의 α 점을 모의방법으로 구하여 한측아래예측한계를 구하였다.

한편 선행연구[3]에서는 정규분포된 모집단으로부터의 단순관측에 대한 예측구간과 한측동시적예측한계를 론의하였으며 선행연구[4]에서는 통계적허용구간에서 β －기대값허용구간이 곧 β 예측구간과 일치한다는데 기초하여 정규분포인 경우에 평균과 분산이 미지일 때 미래단순관측의 량측 β －기대값허용구간을 론의하였다.

선행연구[6]에서는 각각 비파라메터예측으로서의 분포－자유인 경우 예측문제와 정규회귀모형인 경우 믿음직한 예측구간을 구하기 위한 문제를 연구하였다.

이상의 선행연구결과에 기초하여 론문에서는 정규모집단에 대하여 평균과 분산의 가능한 모든 상태에서 미래 m 개관측의 순서통계량과 최소, 최대순서통계량, 표본범위 등의 아래측, 웃측예측한계와 량측예측구간에 대한 정확한 공식들을 제기하였다.

1. 정규모집단으로부터의 미래 m 개관측에 대한 순서통계량의 예측구간

$F(x)$ 와 $f(x)$ 를 각각 모집단분포함수와 밀도함수라고 하자.

련속인 모집단분포함수 $F(x)$ 로부터 크기 m 인 표본 $Y_1, Y_2, \cdots, Y_m$ 의 l 차순서통계량 $Y_{(l)}$ 의 분포함수 $F_{(l)}(x)$ 와 밀도함수 $f_{(l)}(x)$ 는 다음과 같이 주어진다.

$$F_{(l)}(x)=\sum_{i=l}^{m} C_m^i F(x)^i (1-F(x))^{m-i}, \quad f_{(l)}(x)=\frac{1}{B(l,\ m-l+1)} f(x) F^{l-1}(x)(1-F(x))^{m-l}$$

$$F_{(m)}(x)=[F(x)]^m, \quad F_{(1)}(x)=1-[1-F(x)]^m$$

$$f_{(m)}(x)=mf(x)[F(x)]^{m-1}, \quad f_{(1)}(x)=mf(x)[1-F(x)]^{m-1}$$

그리고 표본범위 $R=Y_{(m)}-Y_{(1)}$ 의 분포함수는 $F_R(x)=m\int_{-\infty}^{\infty}[F(u+x)-F(u)]^{m-1} f(u)du$ 이다.

이제 $X_1, X_2, \cdots, X_n$ 이 정규모집단 $N(\mu, \sigma^2)$ 으로부터의 표본(관측된)이라고 하고 $Y_1, Y_2, \cdots, Y_m$ 을 미래 m 개관측으로서 $X_1, X_2, \cdots, X_n$ 과 독립이고 같은 분포 $N(\mu, \sigma^2)$ 에 따른다고 하며 $U_1, U_2, \cdots, U_m$ 을 $Y_1, Y_2, \cdots, Y_m$ 의 표준화된 우연량 즉 $U_i = (Y_i - \mu)\sigma$, $i = 1, \cdots, m$ 이라고 하고 $U_{(1)}, U_{(2)}, \cdots, U_{(m)}$ 을 $U_1, U_2, \cdots, U_m$ 의 순서통계량이라고 하면 $U_{(l)} = (Y_{(l)} - \mu)/\sigma$ 이다. $\mu, \ \sigma^2$ 의 각이한 상태에 대하여 미래 m 개관측 $Y_1, Y_2, \cdots, Y_m$ 의 $Y_{(l)}$ 에 대한 $1-\alpha$ 한측 및 량측예측구간들은 아래와 같이 얻어지는데 얻어지는 예측구간에서 $l = \begin{cases} m/2, & m: \text{짝수} \\ (m+1)/2, & m: \text{홀수} \end{cases}$ 로 놓으면 미래 m 개관측의 표본중위수의 예측구간을 얻을수 있다.

아래의 결과로부터 $\mu, \ \sigma^2$ 이 둘 다 기지인 경우에는 예측구간결정에서 표본값 $X_1, X_2, \cdots, X_n$ 이 리용되지 않는데 그 결과는 다음과 같다.

정리 1 $X_1, X_2, \cdots, X_n$ 이 $N(\mu, \sigma^2)$ (μ , σ^2: 기지)으로부터의 표본이면 $Y_1, Y_2, \cdots, Y_m$ 의 $Y_{(l)}$ 에 대한 $1-\alpha$ 량측예측구간은 $(\mu - k\sigma, \ \mu + k\sigma)$ 이다. 여기서 예측구간인자 $k = k(m, \ 1-\alpha)$ 는 $\sum_{i=l}^{m} C_m^i \{\Phi^i(k)[1-\Phi(k)]^{m-i} - \Phi^i(-k)[1-\Phi(-k)]^{m-i}\} = 1-\alpha$ 를 만족시키는 값이며 Φ 는 $N(0, 1)$ 의 분포함수이다. 즉 $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du$.

또한 $1-\alpha$ 한측예측구간은 $(-\infty, \ \mu + k_1\sigma)$, $(\mu - k_2\sigma, \ +\infty)$ 이다. 여기서 $k_1 = k_1(m, \ 1-\alpha)$ 와 $k_2 = k_2(m, \ 1-\alpha)$ 는 각각 다음의 식들을 만족시키는 값이다.

$$\sum_{i=l}^{m} C_m^i \Phi^i(k_1)[1-\Phi(k_1)]^{m-i} = 1-\alpha \, , \quad \sum_{i=l}^{m} C_m^i \Phi^i(k_2)[1-\Phi(k_2)]^{m-i} = \alpha$$

한측믿을한계와 류사하게 한측예측한계는 앞의 량측예측구간에 대한 식을 적당히 변형하여 얻을수 있다.

이때 $Y_{(l)} > L$ 은 $Y_1, Y_2, \cdots, Y_m$ 의 $m - l + 1$ 개가 L 보다 크다는것과 동등하며 따라서 $1-\alpha$ 웃측예측구간 $(L, \ +\infty)$ 는 관측되지 않은 $m - l + 1$ 개의 우연량들을 포함한다.

반대로 $Y_{(l)} < U$ 는 $Y_1, Y_2, \cdots, Y_m$ 의 적어도 l 개가 U 보다 작다는것과 동등하며 따라서 $1-\alpha$ 아래측예측구간 $(-\infty, \ U)$ 는 적어도 l 개의 관측되지 않은 우연량들을 포함한다.

μ 가 기지이고 σ^2 이 미지인 경우 예측구간을 구하는데 σ^2 의 충분통계량인 $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ 을 리용한다.

정리 2 $X_1, X_2, \cdots, X_n$ 이 정규모집단 $N(\mu, \sigma^2)$ (μ: 기지, σ^2: 미지)으로부터의 표본이면 $Y_1, Y_2, \cdots, Y_m$ 의 $Y_{(l)}$ 에 대한 $1-\alpha$ 량측예측구간은 $(\mu - kS_0, \ \mu + kS_0)$ 이다. 여기서 $k = k(n, \ m, \ 1-\alpha)$ 는 $\int_0^\infty \sum_{i=l}^{m} C_m^i \{\Phi^i(ks_0)[1-\Phi(ks_0)]^{m-i} - \Phi^i(-ks_0)[1-\Phi(-ks_0)]^{m-i}\} h(s_0) ds_0 = 1-\alpha$ 를 만족시키는 값이다. 그리고 $h(s_0)$ 은 $N(0, 1)$ 로부터의 크기 n 인 표본의 표본표준편차의 밀도함수이고 Φ 는 $N(0, 1)$ 의 분포함수이다. 즉

$$h(s_0)=\frac{\nu^{\nu/2}s_0{}^{\nu-1}}{2^{\nu/2-1}\Gamma(\nu/2)}\exp\left(-\frac{\nu s_0^2}{2}\right)\ (s_0\ge 0),\ \Gamma\left(\frac{\nu}{2}\right)=\int_0^\infty x^{\nu/2-1}\exp(-x)dx\ (\nu=n).$$

또한 $Y_{(l)}$에 대한 $1-\alpha$ 한측예측구간은 각각 $(-\infty,\ \mu+k_1S_0)$, $(\mu-k_2S_0,\ +\infty)$이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$와 $k_2=k_2(n,\ m,\ 1-\alpha)$는 각각 다음의 식들을 만족시키는 값이다.

$$\int_0^\infty\left\{\sum_{i=l}^m C_m^i\Phi^i(k_1s_0)[1-\Phi(k_1s_0)]^{m-i}\right\}h(s_0)ds_0=1-\alpha,\ \int_0^\infty\left\{1-\sum_{i=l}^m C_m^i\Phi^i(k_2s_0)[1-\Phi(k_2s_0)]^{m-i}\right\}h(s_0)ds_0=1-\alpha$$

파라메터 μ가 미지이고 σ^2이 기지인 경우에는 예측구간을 구하는데 μ의 충분통계량인 $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$를 리용한다.

정리 3 $X_1,\ \cdots,\ X_n$이 $N(\mu,\ \sigma^2)$ (μ: 미지, σ^2: 기지)으로부터의 표본이면 $Y_1,\ \cdots,\ Y_m$의 $Y_{(l)}$에 대한 $1-\alpha$ 량측예측구간은 $(\overline{X}-k\sigma,\ \overline{X}+k\sigma)$이다. 여기서 $k=k(n,\ m,\ 1-\alpha)$는

$$\int_{-\infty}^\infty\sum_{i=l}^m C_m^i\{\Phi^i(\bar{x}+k)[1-\Phi(\bar{x}+k)]^{m-i}-\Phi^i(\bar{x}-k)[1-\Phi(\bar{x}-k)]^{m-i}\}f(\bar{x})d\bar{x}=1-\alpha$$

를 만족시키는 값이다. 그리고 $f(\bar{x})$는 $N(0,\ 1)$로부터의 표본평균의 밀도함수이고 Φ는 $N(0,\ 1)$의 분포함수이다. 즉 $f(\bar{x})=\sqrt{n/(2\pi)}\exp(-n\bar{x}^2/2),\ -\infty<\bar{x}<\infty$.

또한 $Y_{(l)}$에 대한 $1-\alpha$ 한측예측구간은 각각 $(-\infty,\ \overline{X}+k_1\sigma)$, $(\overline{X}-k_2\sigma,\ +\infty)$이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$와 $k_2=k_2(n,\ m,\ 1-\alpha)$는 각각 다음의 식들을 만족시키는 값들이다.

$$\int_{-\infty}^\infty\left\{\sum_{i=l}^m C_m^i\Phi^i(\overline{X}+k_1)[1-\Phi(\overline{X}+k_1)]^{m-i}\right\}f(\bar{x})d\bar{x}=1-\alpha$$

$$\int_0^\infty\left\{1-\sum_{i=l}^m C_m^i\Phi^i(\overline{X}-k_2)[1-\Phi(\overline{X}-k_2)]^{m-i}\right\}f(\bar{x})d\bar{x}=1-\alpha$$

$f(\bar{x})$는 $N(0,\ 1)$로부터의 표본평균의 밀도함수이고 Φ는 $N(0,\ 1)$의 분포함수이다. 즉

$$f(\bar{x})=\sqrt{n/(2\pi)}\exp(-n\bar{x}^2/2),\ -\infty<\bar{x}<\infty.$$

파라메터 μ와 σ^2이 둘 다 미지인 경우에는 예측구간을 구하는데 μ의 충분통계량 $\overline{X}$와 σ^2의 충분통계량 $S^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2$을 리용한다.

정리 4 $X_1,\ X_2,\ \cdots,\ X_n$이 정규모집단 $N(\mu,\ \sigma^2)$ (μ, σ^2: 미지)으로부터의 표본이면 $Y_1,\ Y_2,\ \cdots,\ Y_m$의 $Y_{(l)}$에 대한 $1-\alpha$ 량측예측구간은 $(\overline{X}-kS,\ \overline{X}+kS)$이다. 여기서 $k=k(n,\ m,\ 1-\alpha)$는 다음의 방정식을 만족시키는 값이다.

$$\int_0^\infty g(s)\int_{-\infty}^\infty\sum_{i=l}^m C_m^i\{\Phi^i(\bar{x}+ks)[1-\Phi(\bar{x}+ks)]^{m-i}-\Phi^i(\bar{x}-ks)[1-\Phi(\bar{x}-ks)]^{m-i}\}f(\bar{x})d\bar{x}ds=1-\alpha$$

그리고 $f(\bar{x})$, $g(s)$는 각각 $N(0,\ 1)$로부터의 표본평균과 표본표준편차 S의 밀도함수이고 Φ는 $N(0,\ 1)$의 분포함수이다. 즉 $g(s)=\frac{\nu^{\nu/2}s^{\nu-1}}{2^{\nu/2-1}\Gamma(\nu/2)}\exp\left(-\frac{\nu s^2}{2}\right),\ s\ge 0,\quad \nu=n-1.$

또한 $Y_{(l)}$ 에 대한 $1-\alpha$ 한측예측구간은 각각 $(-\infty,\ \overline{X}+k_1S)$, $(\overline{X}-k_2S,\ +\infty)$ 이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$ 와 $k_2=k_2(n,\ m,\ 1-\alpha)$ 는 각각 다음의 식을 만족시키는 값이다.

$$\int_0^\infty g(s)\int_{-\infty}^\infty\left\{\sum_{i=l}^m C_m^i\Phi^i(\overline{X}+k_1s)[1-\Phi(\overline{X}+k_1s)]^{m-i}\right\}f(\overline{x})d\overline{x}ds=1-\alpha$$

$$\int_0^\infty g(s)\int_{-\infty}^\infty\left\{1-\sum_{i=l}^m C_m^i\Phi^i(\overline{X}-k_2s)[1-\Phi(\overline{X}-k_2s)]^{m-i}\right\}f(\overline{x})d\overline{x}ds=1-\alpha$$

2. 정규모집단에 대한 미래 최소, 최대순서통계량들의 한측예측한계

$X_1, X_2, \cdots, X_n$ 이 정규모집단 $N(\mu,\ \sigma^2)$ 으로부터의 표본이고 $Y_1, Y_2, \cdots, Y_m$ 이 미래 m 개관측으로서 $X_1, X_2, \cdots, X_n$ 과 독립이고 같은 분포 $N(\mu,\ \sigma^2)$ 에 따른다고 하자.

정리 5 $X_1, X_2, \cdots, X_n$ 이 $N(\mu,\ \sigma^2)$ 으로부터의 표본이면 미래 m 개관측 $Y_1, Y_2, \cdots, Y_m$ 의 최소(최대)순서통계량 $Y_{(1)}$ $(Y_{(m)})$에 대한 $1-\alpha$ 한측예측한계는 다음과 같다.

① μ, σ^2 이 기지인 경우 $(-\infty,\ \mu+k_1\sigma)$, $(\mu-k_2\sigma,\ +\infty)$ 이다. 여기서 $k_1=k_1(m,\ 1-\alpha)$ 와 $k_2=k_2(m,\ 1-\alpha)$ 는 각각 다음의 식을 만족시키는 값이다.

$$[1-\Phi(k_1)]^m=\alpha,\ [1-\Phi(k_2)]^m=1-\alpha\quad(\Phi^m(k_1)=1-\alpha,\ \Phi^m(-k_2)=\alpha)$$

② μ 가 기지, σ^2 이 미지인 경우 $(-\infty,\ \mu+k_1s_0)$, $(\mu-k_2s_0,\ +\infty)$ 이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$ 와 $k_2=k_2(n,\ m,\ 1-\alpha)$ 는 각각 다음의 식들을 만족시키는 값이다.

$$\int_0^\infty\{1-[1-\Phi(k_1s_0)]^m\}h(s_0)ds_0=1-\alpha,\ \int_0^\infty\Phi^m(k_2s_0)h(s_0)ds_0=1-\alpha$$

$$\left(\int_0^\infty\Phi^m(k_1s_0)h(s_0)ds_0=1-\alpha,\ \int_0^\infty[1-\Phi^m(-k_2s_0)]h(s_0)ds_0=1-\alpha\right)$$

③ μ 가 미지, σ^2 이 기지인 경우 $(-\infty,\ \overline{X}+k_1\sigma)$, $(\overline{X}-k_2\sigma,\ +\infty)$ 이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$ 와 $k_2=k_2(n,\ m,\ 1-\alpha)$ 는 각각 다음의 식을 만족시키는 값이다.

$$\int_{-\infty}^\infty\{1-[1-\Phi(\overline{x}+k_1)]^m\}f(\overline{x})d\overline{x}=1-\alpha,\ \int_{-\infty}^\infty[1-\Phi(\overline{x}-k_2)]^m f(\overline{x})d\overline{x}=1-\alpha$$

$$\left(\int_{-\infty}^\infty\Phi^m(\overline{x}+k_1)f(\overline{x})d\overline{x}=1-\alpha,\ \int_{-\infty}^\infty[1-\Phi^m(\overline{x}-k_2)]f(\overline{x})d\overline{x}=1-\alpha\right)$$

④ μ, σ^2 이 미지인 경우 $(-\infty,\ \overline{X}+k_1s)$, $(\overline{X}-k_2s,\ +\infty)$ 이다. 여기서 $k_1=k_1(n,\ m,\ 1-\alpha)$ 와 $k_2=k_2(n,\ m,\ 1-\alpha)$ 는 각각 다음의 식을 만족시키는 값이다.

$$\int_{-\infty}^\infty g(s)\int_0^\infty\{1-[1-\Phi(\overline{x}+k_1s)]^m\}f(\overline{x})d\overline{x}ds=1-\alpha,\ \int_{-\infty}^\infty g(s)\int_0^\infty\{1-[1-\Phi(\overline{x}-k_2s)]^m\}f(\overline{x})d\overline{x}ds=1-\alpha$$

$$\left(\int_{-\infty}^\infty g(s)\int_0^\infty\Phi^m(\overline{x}+k_1s)f(\overline{x})d\overline{x}ds=1-\alpha,\ \int_{-\infty}^\infty g(s)\int_0^\infty[1-\Phi^m(\overline{x}-k_2s)]f(\overline{x})d\overline{x}ds=1-\alpha\right)$$

3. 정규모집단에 대한 미래표본범위의 예측구간

$N(\mu,\ \sigma^2)$ 으로부터의 크기 n 인 표본 $X_1, X_2, \cdots, X_n$ 에 기초하여 같은 분포에 따르는 미래 m 개관측 $Y_1, Y_2, \cdots, Y_m$ 에 대한 표본범위 R 의 $1-\alpha$ 예측구간을 구하기로 한다.

$Y_1, Y_2, \cdots, Y_m$ 이 $N(\mu,\ \sigma^2)$ 으로부터의 독립인 표본일 때 표본범위 $R = Y_{(m)} - Y_{(1)}$ 에 대하여 수학적기대값과 분산은 각각 $\mathrm{E}R = d_2\sigma$, $\mathrm{Var}R = d_3^2\sigma^2$ ($d_2,\ d_3$ 은 m 에 따르는 수값)으로서 R 의 예측은 σ 에 대한 충분통계량인 S 와 관련된다.

$X_1, X_2, \cdots, X_n$ 이 정규모집단 $N(\mu,\ \sigma^2)$ 으로부터의 표본이고 $Y_1, Y_2, \cdots, Y_m$ 을 미래 m 개관측으로서 $X_1, X_2, \cdots, X_n$ 과 독립이고 같은 분포 $N(\mu,\ \sigma^2)$ 에 따른다고 한다.

정리 6 $X_1, X_2, \cdots, X_n$ 이 $N(\mu,\ \sigma^2)$ 으로부터의 표본이고 이 표본에 대한 표준편차를 S 라고 하면 미래 m 개관측 $Y_1, Y_2, \cdots, Y_m$ 의 표본범위 $R = Y_{(m)} - Y_{(1)}$ 의 $1-\alpha$ 예측구간은 $(k_1S,\ k_2S)$ 와 같다. 여기서 $k_1,\ k_2$ 는 각각 다음의 식을 만족시키는 값이다.

$$\int_0^\infty g(s)\left[\int_{-\infty}^\infty m[\Phi(u)-\Phi(u-k_2s)]^{m-1}\phi(u)du\right]ds = \frac{1-\alpha}{2},\quad \int_0^\infty g(s)\left[\int_{-\infty}^\infty m[\Phi(u)-\Phi(u-k_1s)]^{m-1}\phi(u)du\right]ds = \frac{1-\alpha}{2}$$

여기서 $\phi(x)$ 는 표준정규밀도함수이다.

참 고 문 헌

[1] 한광룡 등; 통계적구간, 고등교육도서출판사, 23∼193, 주체104(2015).

[2] G. J. Hahn; Statistical Intervals: A Guide for Pratitioners, Wiley, 56∼387, 1991.

[3] D. K. Bhaumik; The Indian Journal of Statistics, B 70, 248, 2008.

[4] K. Krishnamoorthy; Statistical Tolerance Regions, Wiley, 89∼461, 2009.

[5] C. M. Wang et al.; Journal Statistical Planning and Inference, 142, 1980, 2012.

[6] M. G. Hamed et al.; arXiv: 1603.05587v4[Stat.ME], 1, Apr, 2016.

주체105(2016)년 7월 5일 원고접수

Prediction Intervals for Ordered Statistic and Sample Range of Future m Observations from a Normal Population

An Jong Hwa, Han Kwang Ryong

We obtained one-sided prediction limits and two-sided prediction intervals for ordered statistic and the smallest, largest ordered statistic and sample range of future m observations from a normal population.

Key words: ordered statistic, prediction interval