The Korean Journal of Applied Statistics (2020) DOI: https://doi.org/10.5351/KJAS.2020.33.4.357
33(4), 357–364

A brief study on the geometric mean

In-Kwon Yeo[a,1]

[a]Department of Statistics, Sookmyung Women's University

(Received April 6, 2020; Revised April 23, 2020; Accepted April 23, 2020)

Abstract

We review the characteristics of a geometric mean and statistical inferences based on geometric means. We also show that the statistical results obtained by the logarithmic transform and back-transformation are related to geometric means and explain how to interpret the results produced in this process.

Keywords: geometric mean, logarithmic transformation, exponential function

1. 서론

통계학과 학부과정에서의 기술통계와 통계적 추론은 대부분 산술평균을 중심으로 강의가 이루어진다. 일반적인 선형회귀분석이나 분산분석에서도 결국 산술평균에 대한 통계적 추론을 다루고 있다. 평균기반 자료분석에서는 자료가 등분산성이나 정규성 가정을 심각하게 위배한 경우, 이를 해결하는 방법 중 하나로 로그변환과 같은 변수 변환을 한 후 분석하는 방법을 소개하고 있다. 문제는 변수변환 후에 분석결과를 잘못 이해하고 활용하는 경우가 많다는 것이다. 이 소고는 실증분석에서 흔히 볼 수 있는 로그변환된 자료의 분석에 대한 잘못된 해석을 바로 잡고자 준비하였다. 로그변환은 기하평균과 밀접한 관련이 있어 이 소고에서는 기하평균에 대한 여러 성질을 알아보고 변환-역변환 방법을 통한 통계적 추론에서 주의해야 할 점에 대해 알아보고자 한다.

2. 기하평균의 성질

금융상품의 수익률에 대한 분석에서 P_t를 t시점에서의 상품 가격이라고 할 때 단순 수익률(simple return)을 $R_t = (P_t - P_{t-1})/P_{t-1}$로 정의하지만 분석의 편의성을 위해 실제 자료분석에서는 로그 수익률(log return) $r_t = \log(P_t/P_{t-1})$을 사용하고 있다. 일정 기간 동안의 평균수익률을 이해하기 위해 간단한 예를 들어 보자. 어떤 시점에서 100을 투자한 후 1시점에서는 10% 수익이 발생하고 2시점에서는 10% 손실이 발생했다고 했을 때, 2시점에서의 가격은 $100 \times (1 + 0.1) \times (1 - 0.1) = 99$가 된다. 즉, 100을 투자해서 99가 되었으며 2시점 동안 1%만큼 손해를 봤으며 시점별 평균수익률은 -0.5%로 일반인을 오해할 수 있다. 하지만 가격 산출과정에서 시점별 평균수익률의 의미는 $100 \times (1 + 0.1) \times (1 - 0.1) = 100 \times (1 + a) \times (1 + a)$가 성립하는 a값을 의미하며 $(1 + a)^2 = (1 + 0.1)(1 - 0.1)$이 되고 $a = \sqrt{1.1 \times 0.9} - 1 = -0.005016$가 된다. 이를 좀더 일반화하기 위해 $x_t = P_t/P_{t-1}$, 즉 P_t가 P_{t-1}의

[1]Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Gu, Seoul 04310, Korea. E-mail: inkwon@sookmyung.ac.kr

몇 배인지를 나타내는 값이고 $x_G = (1+a)$를 평균 배수라고 하면 위의 예제는 $x_G^2 = x_1 \times x_2$가 된다. 만약 투자를 n시점까지 진행을 했다면 곱해지는 항수가 n만큼 늘어나고 x_G의 승수 또한 n만큼 늘어 나게 된다.

$$x_G^n = x_1 \times \cdots \times x_n = \frac{P_n}{P_0}.$$

결론적으로 다음과 같은 관계식이 유도된다.

$$x_G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \left(\frac{P_n}{P_0}\right)^{\frac{1}{n}}.$$

자료의 측정 단위가 배수의 형태로 표시되는 경우에는 중심위치의 측도로 산술평균을 사용하는 것은 자료의 특성을 제대로 반영하지 못한 것으로 볼 수 있다. 이는 단위가 %P가 아닌 % 형태로 되어 있는 자료의 경우도 마찬가지라고 할 수 있다.

예제 2.1: e-나라지표에 의하면 1인단 실질 국민총소득을 원화로 표시했을 때 1988년 972만원, 2018년 3493만원인 것으로 조사되었으며 이를 이용하여 30년 동안 1인당 실직 국민총소득의 연평균 증가율을 계산하면

$$\left(\frac{3493}{972}\right)^{\frac{1}{30}} = 1.0436$$

으로 4.36%인 것으로 나타났다. 추가적으로 1998년 1인당 실질 국민총소득은 1591만원으로 1988년부터 1998년 10년 동안 연평균 증가율과 1998년부터 2018년 20년 동안 연평균 증가율은 아래와 같은 식을 이용하여 각각 5.05%와 4.01%인 것을 보일 수 있다.

$$\left(\frac{1591}{972}\right)^{\frac{1}{10}} = 1.0505, \quad \left(\frac{3493}{1591}\right)^{\frac{1}{20}} = 1.0401$$

이 연평균 배수를 이용하여 1988년부터 2018년까지의 연평균 배수를 가중기하평균을 이용하여 구할 수 있다.

$$\left(1.0505^{10} \times 1.0401^{20}\right)^{\frac{1}{30}} = 1.0436.$$

기하평균을 계산함에 있어 양수 자료가 계속 곱해지는 과정에서 컴퓨터가 처리할 수 있는 수의 한계를 초과할 수 있어 일반적으로 아래와 같이 로그 변환한 자료의 평균을 계산한 후 지수함수를 적용하여 계산한다.

$$\bar{x}_G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(x_i)\right), \quad x_i > 0.$$

예제 2.1에서 언급된 가중기하평균의 경우 w_i를 i번째 관측값의 가중치라고 하면 다음과 같이 정의 될 수 있다.

$$\bar{x}_{WG} = \left(\prod_{i=1}^{n} x_i^{w_i}\right)^{\frac{1}{W}} = \exp\left(\frac{1}{W}\sum_{i=1}^{n} w_i \log(x_i)\right), \quad W = \sum_{i=1}^{n} w_i.$$

즉, 기하평균은 로그변환한 자료를 이용하여 산술평균을 구하고 역함수인 지수함수에 대입한 것이다. 이는 통계분석에서 로그변환 후 분석결과를 다시 역변환하여 원자료의 척도를 바꾸어 해석하는 과정과 유사하다. 여기서 주목해야 할 것은 만약 로그변환 후 추론의 대상이 산술평균이었다면 역변환의 결과는 산술평균이 아니라 기하평균이라는 것이다.

3. 모집단의 기하평균

로그함수는 오목(concave)함수이기 때문에 Jensen의 부등식을 이용하면 기하평균이 산술평균보다 항상 작거나 같다는 것을 보일 수 있다 (Casella와 Berger, 1990, 예제 4.7.3). 표본에 대한 기하평균 공식은 많이 알려져 있으나 모집단의 기하평균은 상대적으로 알려져 있지 않다. 이를 유도하기 위해서 통계적 확률 개념을 이용하여 n이 계속 커지면 표본평균 $\bar{x}$가 모평균 μ로 수렴하는 것을 볼 수 있다.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \mu = E(X),$$

여기서 $\lim_{n\to\infty} f_n = f$ 형태는 수학에서의 수렴이기 보다는 큰 수의 법칙과 같이 확률적으로 f_n이 f로 수렴한다는 것을 표시한 것이다. 이론적으로는 이것이 성립하기 위해서는 $E(X)$가 존재해야 한다. 모집단의 산술평균과의 구분을 위해 모집단의 기하평균을 μ_G라고 표시하자. 위의 통계적 확률의 개념을 이용하면

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log(x_i) = E(\log(X))$$

가 되고 $E(\log(X)) < \infty$라는 조건 하에서 exp는 연속함수이므로 이 μ_G를 기댓값의 형태로 표시하면 다음과 같이 정의된다는 것을 보일 수 있다.

$$\begin{aligned}\lim_{n\to\infty} \bar{x}_G &= \lim_{n\to\infty} \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right)\\ &= \exp\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right) = \exp\left(E(\log(X))\right) = \mu_G,\end{aligned}$$

여기서 강조하고 싶은 것은 어떤 이유에서 자료를 로그변환하고 산술평균과 관련된 통계적 추론을 했다면 이는 원자료의 척도에서의 산술평균이 아닌 기하평균에 대한 추론이라는 것이다.

4. 기하평균 추론

분포의 중심위치 추론과정에서 확률표본 $X_1, \ldots, X_n$에 대해 $Y_i = \log(X_i)$를 구하고 Y_i의 분포가 다음과 같다고 가정하자.

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N\left(\mu_Y, \sigma_Y^2\right).$$

이것은 $X_i = \exp(Y_i)$는 로그정규분포(log-normal distribution)을 따른다는 것을 의미한다. Y들의 표본평균을 $\bar{Y}$라고 하면 모평균 μ_Y에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$(L,U) = \left(\bar{Y} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}},\ \bar{Y} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right),$$

여기서 $t_{\alpha/2,n-1}$는 자유도가 $n-1$인 t-분포에서 $1-\alpha/2$에 해당하는 분위수를 의미하고 S는 $Y = \log(X)$들의 표본표준편차를 의미한다. 변환자료의 산술평균을 지수함수로 역변환하면 $\exp(\bar{Y}) = \bar{X}_G$가 되고 이는 $E(X)$가 아닌 $\exp(\mu_Y) = \mu_G$를 추론하기 위한 통계량이며 μ_G에 대한 신뢰구간 또한 지수변환을 통해 다음과 같이 구한다.

$$(\exp(L), \exp(U)) = \left(\frac{\bar{X}_G}{\exp\left(t_{\frac{\alpha}{2},n-1}S/\sqrt{n}\right)},\ \bar{X}_G \exp\left(t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right)\right).$$

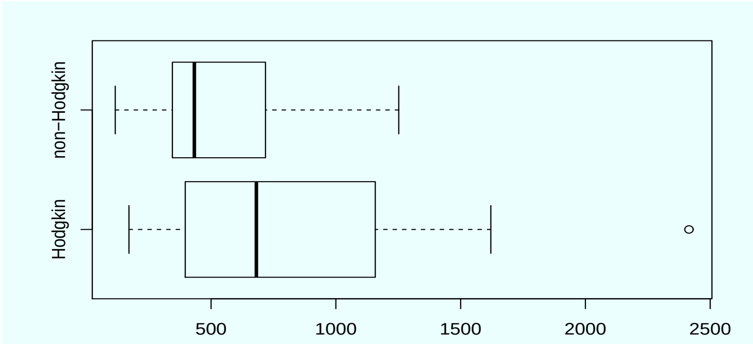

Figure 4.1. Boxplot of T_4 cell counts/mm^3 in patients with and without Hodgkin's disease.

두 모집단 A와 B의 중심위치를 비교하는 문제에서 μ_A와 μ_B를 각 모집단의 중심위치라고 하자. 두 중심위치가 같다, 즉 $\mu^A = \mu^B$는 ① $\mu^A - \mu^B = 0$ 또는 ② $\mu^A/\mu^B = 1$로 표시할 수 있다. 통계분석에서 흔히 실수하는 것 중 하나가 평균의 비교를 항상 ①과 같은 형태로 생각하는 것이다. 만약 자료를 로그 변환을 한 후 평균의 차로 중심위치를 비교했다고 하면

$$\bar{Y}^A - \bar{Y}^B = \frac{1}{m}\sum_{i=1}^{m}\log\left(X_i^A\right) - \frac{1}{n}\sum_{i=1}^{n}\log\left(X_i^B\right) = \log\left(\bar{X}_G^A\right) - \log\left(\bar{X}_G^B\right).$$

이것은 $\log(\mu_G^A) - \log(\mu_G^B) = 0$, 즉 $\mu_G^A/\mu_G^B = 1$인지를 확인하기 위한 통계량으로 두 기하평균이 같은지를 비교하기 위한 것이다. 즉, 기하평균의 비교는 차로 표시하는 것이 아니라 비로 표시해야 함을 의미한다. 만약 두 모집단의 분산이 동일하다면 $\log(\mu_G^A) - \log(\mu_G^B)$에 대한 신뢰구간은 다음과 같다.

$$\left(\bar{Y}^A - \bar{Y}^B - t_{\frac{\alpha}{2},m+n-2}S_p\sqrt{\frac{1}{m}+\frac{1}{n}},\ \bar{Y}^A - \bar{Y}^B + t_{\frac{\alpha}{2},m+n-2}S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right).$$

이를 기반으로 원자료의 척도에서 평균을 비교한다면 위 구간의 경계값을 지수함수에 대입해야 하며 결론적으로 기하평균의 비 μ_G^A/μ_G^B에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이 구할 수 있다.

$$\left(\frac{\bar{X}_G^A/\bar{X}_G^B}{\exp\left(t_{\frac{\alpha}{2},m+n-2}S_p\sqrt{1/m+1/n}\right)},\ \frac{\bar{X}_G^A}{\bar{X}_G^B}\exp\left(t_{\frac{\alpha}{2},m+n-2}S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right)\right). \tag{4.1}$$

이를 좀 더 일반화하여 분산분석에도 적용할 수 있는데 로그 변환한 자료의 평균 비교는 수준 간 평균의 차로 해석할 수 있으나 만약 역변환을 한 후 결과를 해석 한다면 수준 간의 기하평균의 비로 해석해야 한다는 것을 의미한다.

예제 4.1: Figure 4.1은 악성림프종의 하나인 Hodgkin병에서 차도가 있는 환자 20명과 파종성 악성종양(non-Hodgkin)에서 차도가 있는 환자 20명의 1mm^3 혈액 샘플에 있는 T_4 세포의 수에 대한 상자그림으로 두 자료 모두 양의 왜도를 가지는 것을 볼 수 있다 (Shapiro 등, 1987). 두 자료에 대해 자연로그 변환한 후 Shapiro-Wilk 검정을 실시한 결과 p-값이 Hodgkin은 0.8117, non-Hodgkin은 0.4369로 정규성을 만족했고 등분산성 검정에서도 p-값이 0.6223으로 위의 식 (4.1)을 적용하여 두 기하평균의 비에 대한 신뢰구간을 구할 수 있다. 신뢰구간을 구하기 위한 주요 통계값은 다음과 같으며

자료	표본크기	로그평균($\bar{Y}$)	기하평균($\bar{X}_G$)
Hodgkin	20	6.487	656.51
Non-Hodgkin	20	6.089	440.82

두 기하평균의 비는 656.51/440.82 = 1.4893로 Hodgkin의 T_4 수가 non-Hodgkin의 T_4 수의 1.49배 정도 많은 것으로 나타났으며 로그변환한 자료의 합동표준편차가 $s_p = 0.671$로 Hodgkin과 non-Hodgkin의 기하평균에 대한 95% 신뢰구간은 다음과 같이 구할 수 있다.

$$\left(\frac{1.4893}{\exp\left(2.024 \times 0.671\sqrt{1/20+1/20}\right)},\ 1.4893\exp\left(2.024 \times 0.671\sqrt{\frac{1}{20}+\frac{1}{20}}\right)\right) = (0.9692, 2.2884)$$

이 신뢰구간이 1을 포함하기 때문에 두 기하평균 간에는 유의한 차이가 없는 것으로 볼 수 있다.

로그변환한 자료를 이용하여 $E(X) = \mu$에 대한 신뢰구간을 계산하는 방법은 여러 가지가 있으며 관련 식은 Zhou와 Gau (1997)을 참조할 수 있다. 이들 방법 중 Cox 방법은 로그정규분포에서 $\mu = \exp(\mu_Y + \sigma_Y^2/2)$로 표시되기 때문에 $\log(\mu)$을 $\bar{Y} + S^2/2$으로 추정하고 이 추정량의 분산을 $S^2/n + S^4/(2(n-1))$으로 추정하여 다음과 같이 구한다.

$$\left(\exp\left(\bar{Y}+\frac{S^2}{2}-z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}+\frac{S^4}{2(n-1)}}\right),\ \exp\left(\bar{Y}+\frac{S^2}{2}+z_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}+\frac{S^4}{2(n-1)}}\right)\right),$$

여기서 z_α는 표준정규분포에서 $(1-\alpha)$분위수를 의미한다. 동일한 방법을 이용하여 이표본 평균비교하는 방법은 Zhou 등 (1997)에 자세히 설명되어 있다.

5. 선형모형에서의 기하평균

분산분석이나 회귀분석의 통계적 추론에서 추정값의 변동성의 측도는 모두 MSE을 기반으로 하고 있으며 이 MSE는 등분산성 가정 하에서 유도되기 때문에 다른 가정보다 중요하게 취급된다. 등분산성 가정을 심각하게 위반하는 경우 반응변수를 변환하여 등분산성 조건을 가능한 충족하게 하는 방법이 오래전부터 연구되어 왔다. 대표적인 방법이 Bartlett (1947)의 분산안정화변환(variance stabilizing transformation)과 Box-Cox (1964) 변환이 있다. 분산안정화변화에서는 표준편차가 평균의 상수배로 표시되는 경우 로그변환을 추천하고 있다. 이는 잔차그림에서 평균에 해당되는 예측값이 커질 때 잔차 폭의 경계가 직선형태로 넓어지는 경우에 로그 변환한 반응변수로 새로 분석하면 잔차의 폭이 수평 경계 내에서 고르게 분포될 가능성이 있다는 것이다. Box-Cox의 경우에는 등분산성, 정규성, 선형성을 모두 개선시키는 것을 목적으로 하고 있으며 최대가능도추정으로 변환모수를 추정했을 때 그 값이 0 근처에 있으면 로그변환이 이들 가정을 최대한 충족하게 만든다는 것을 의미한다.

실제 자료분석에서는 다른 변환보다는 로그변환이 많이 사용되고 있으며 로그변환을 한 후 모형을 적합시키고 지수함수를 통해 역변환화여 원자료의 척도에 대한 결과로 해석하는 경우가 많다. 이렇게 변환 후 다시 역변환하는 방법을 변환-역변환(transform and back-transform) 또는 변환-재변환(transform and retransform) 방법이라고 하는데 이 방법은 분위수에 대한 추론에서는 문제가 발생하지 않으나 평균에 대한 추론에서는 문제가 발생한다. 회귀분석에서 다음과 같이 로그변환한 반응변수로 모형식을 설정한 후

$$\log(Y_i) = Z_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

최소제곱법으로 회귀계수를 구한 후 예측값을 구하면 다음과 같다.

$$\hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip},$$

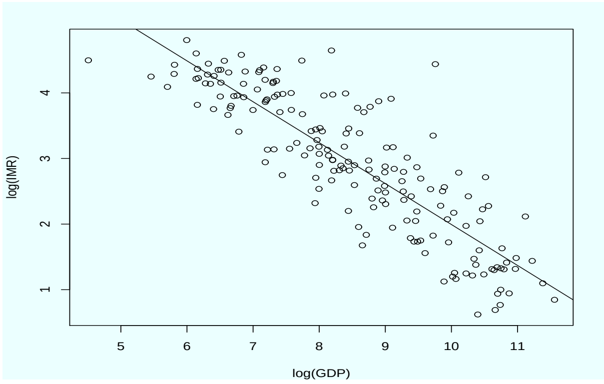

Figure 5.1. Regression of the logarithm of the infant mortality rate on the log of GDP/capita.

여기서 중요한 것은 $\hat{z}_i$는 $E(Z_i)$의 추정값이고 Y의 관점에서는 $E(\log(Y_i))$의 추정값인 것이다. 임의의 비선형 함수 $g(\ )$에 대해 $g(E(X)) \neq E(g(X))$이므로, 만약 Y의 척도로 결과를 해석하기 위해 지수함수를 이용하여 $\exp(\hat{z}_i)$를 구했다면 이는 $E(Y_i)$에 대한 추정값이 아니라 $\exp(E(\log(Y_i)))$의 추정값을 구한 것이다. 즉, $\exp(\hat{z}_i)$는 산술평균인 $E(Y_i)$의 추정값이 아닌 기하평균인 $\exp(E(\log(Y_i)))$의 추정값인 것이다. 추정된 회귀계수 $\hat{\beta}_j$는 j번째 설명변수 x_j를 한 단위 증가함에 따라 반응변수 Y의 기하평균이 이전보다 $\exp(\hat{\beta}_j)$배가 된다는 것을 의미한다. 로그변환이 일반화되어 있어 경제 논문이나 유명 블로그에서는 이에 대한 해석으로 "반응변수가 몇 % 증감하였다"고 비슷한 어감으로 설명하고 있으나 반응변수의 무엇이 증감했는지에 대한 구체적 설명이 추가된 논문은 거의 찾아 볼 수 없다. 이제 통계전문가라면 로그변환-역변환의 결과는 반응변수의 산술평균이 아닌 기하평균의 배수에 대한 분석이라고 설명할 수 있어야 한다고 생각된다.

예제 5.1: Figure 5.1은 UNdata에서 제공한 181개 국가 및 지역의 2010년 GDP/capita (GDP)와 출생아 1,000명 기준 유아사망률(IMR)을 각각 자연로그 변환하여 표시한 것으로 그림에서 선으로 표시된 추정 회귀식은 다음과 같다.

$$\log(\mathrm{IMR}) = 8.23144 - 0.6238 \log(\mathrm{GDP}).$$

통상적으로 GDP와 같은 자료는 설명변수와 반응변수 관계없이 로그변환하여 모형에 적용하는데 위의 모형과 같이 설명변수에 포함되어 있는 경우에 GDP값의 조정은 얼마 증가 또는 감소 보다는 몇 % 또는 몇 배 증가 또는 감소로 이루어지는 것이 더 적절하다. 위의 추정식을 통해 GDP/capota가 10% 증가되면 $\exp(-0.6238 * \log(1.1)) = 0.9423$으로 유아사망률의 기하평균은 5.77% 정도 감소하고 10% 감소하면 $\exp(-0.6238 * \log(0.9)) = 1.0679$으로 유아사망률의 기하평균이 6.79% 정도 증가하는 것으로 해석할 수 있다.

만약 변환-역변환 방법을 적용했을 때 산술평균에 대한 추정값을 구하려면 Duan (1983)의 Smearing 추정법을 적용할 수 있다.

6. 결론

지금까지 표본의 기하평균과 모집단의 기하평균, 기하평균에 대한 추론과 선형모형에서의 기하평균에 대해 로그변환과 지수 역변환을 이용하여 어떻게 해석하고 활용해야 하는지에 대해 설명하였다. 외부

컨설팅 과정에서 적절하지 않은 분석방법 사용되고 결과도 잘못 해석하고 있는 것을 여러 기관, 심지어 국가정책을 세우는 국책연구기관에서도 목격했으며 해당 업무 연구원도 기존에 그렇게 사용해 왔고 누구도 지적하지 않아 그대로 사용하고 있다는 답변을 들었는데 이런 일이 앞으로는 발생하지 않도록 하는 것이 필요하다는 생각에 이 소고를 쓰게 되었다. 새로운 통계기법의 개발만큼 잘 만들어진 기존 통계기법을 제대로 활용하는 것도 중요하다고 생각되며 잘못 사용되고 있는 유사한 사례의 공유와 개선 제안 등이 이어질 것을 기대한다.

References

Barlett, M. S. (1947). The use of transformations, *Biometrics*, **3**, 39–52.

Box, G. E. P and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26**, 211–252.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*, Thomson Information/Publishing Group.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method, *Journal of the American Statistical Association*, **78**, 605–610.

Shapiro, C. M., Beckmann, E., Christiansen, N., Bitran, J. D., Kozloff, M., Billings, A. A., and Telfer, M. C. (1987). Immunologic status of patients in remission from Hodgkin's disease and disseminated malignancies, *The American Journal of the Medical Sciences*, **293**, 366–370.

Zhou, X. H. and Gao, S. (1997). Confidence intervals for the log-normal mean, *Statistics in Medicine*, **16**, 783–790.

Zhou, X. H., Gao, S., and Hui, S. L. (1997). Methods for comparing the means of two independent log-normal samples, *Biometrics*, **53**, 1129–1135.

기하평균에 대한 소고

여인권[a,1]

[a]숙명여자대학교 통계학과

(2020년 4월 6일 접수, 2020년 4월 23일 수정, 2020년 4월 23일 채택)

요 약

이 소고에서는 기하평균의 성질과 기하평균과 관련된 통계적 추론에 대한 알아 본다. 로그변환-역변환을 통해 얻어진 통계적 추론 결과가 기하평균과 관련이 있다는 것을 보이고 이 과정에서 유도된 결과를 어떻게 해석해야 하는지를 설명한다.

주요용어: 기하평균, 로그변환, 지수함수

[1](04310) 서울시 용산구 청파로47길 100, 숙명여대 통계학과. E-mail: inkwon@sookmyung.ac.kr