CONFIDENCE INTERVAL ESTIMATION FOR VALUE OF TIME*

by Hisa MORISUGI*, Jane ROMERO** and Takayuki MORIGUCHI ***

1. Introduction

The subjective value of time is the marginal rate of substitution between travel time and travel cost. It is commonly referred to as willingness to pay of an individual for savings in travel time. In practice, it is derived normally from discrete choice models based on random-utility theory (Ben-Akiva and Lerman, 1985). The resulting willingness to pay value is a point estimate from the mean of travel time divided by the mean of the travel cost, as if the pseudo mean of the distribution of the ratio X/Y would be just the ratio of the means of μ_1/μ_2. In addition, the single value of the subjective value of time is too crude a summary considering the vast sample size.

Garrido and Ortúzar (1993) proposed replacing the single value by the construction of a confidence interval given a certain level of confidence. This allows the estimation of lower and upper limits, which is important in the sensitivity analyses of project evaluation. Further, Armstrong, Garrido and Ortúzar (2001) proposed two methods, the asymptotic t-test and likelihood ratio test, to make statistical inference on the ratio without the direct use of the associated probability density function since they considered the probability distribution for the ratio between two normally distributed variables as unknown a priori. However, it is a well-solved problem albeit a messy and complex solution.

Thus, the basic objectives of this paper are to first, discuss the theory of ratio of normal and its applicability to bind the value of time and then to discuss the t-test method as an alternative estimation method to build the confidence interval for the ratio. Section 2 will discuss further the theory of ratio of normal variables and the possible forms and shapes of the distribution. In Section 3 is the discussion on the methods of building the confidence interval, the direct substitution method and t-test method. The application to value of time and comparison of the methods are in Section 4. Section 5 summarizes the conclusions.

*Keywords: value of time, confidence interval, ratio of normal random variables, t-test method

*Fellow of JSCE, Ph.D., Professor
Graduate School of Information Sciences, Tohoku University
(Aoba 06, Aoba, Sendai, Japan,TEL:22-217-7501,FAX:22-217-7500)
**Graduate Student, Graduate School of Information Sciences, Tohoku University
(Aoba 06, Aoba, Sendai, Japan,TEL:22-217-7501,FAX:22-217-7500)
***Member of JSCE, SELCO Home Corp.
(2-1-14 Kamisugi, Aoba, Sendai, Japan,TEL:22-224-1110,FAX:22-511-0303)

(1) Definition of value of time (VOT) and the assumed modal split model

Value of time is defined as the change in travel cost relative to change in travel time with the utility level kept constant. The subjective value of time is given as

$$q = -\frac{dP_i}{dT_i}\bigg|_{U_{i=const}} = \frac{\frac{dy}{dT_i}}{\frac{dy}{dP_i}} = \frac{a_2}{a_1} \tag{1.1}$$

where: P_i = travel cost
T_i = travel time

Travel cost and travel time are variables of a general multivariate normal population. Assuming an aggregate Logit model, an individual has the following choice of modes (Eqns. 1.2, 1.3) given the utility function as Eqn. 1.4.

$$\Pi_1 = \frac{e^{U_1}}{e^{U_1} + e^{U_2}} \tag{1.2}$$

$$\Pi_2 = \frac{e^{U_2}}{e^{U_1} + e^{U_2}} \tag{1.3}$$

$$U_i = a_{0i} + a_1 p_i + a_2 t_i \tag{1.4}$$

where: Π_i : share of mode *i*
U_i : utility function of mode *i*
$\alpha_{0i}, \alpha_1, \alpha_2$: parameters
p_i = travel cost of mode *i*
t_i = travel time of mode *i*

The linear willingness to pay function is given by Eqn. 1.5 as,

$$y_{ij} = \ln \frac{\Pi_{1j}}{1-\Pi_{1j}} \qquad j = 1\ldots n,$$

$$y_{ij} = a_{01} - a_{02} + a_1(p_{1j} - p_{2j}) + a_2(t_{1j} - t_{2j}) + e_{1j} \tag{1.5}$$

Let $a_0 = (a_{01} - a_{02})$, $P_i = (p_1 - p_2)$, $T_i = (t_1 - t_2)$ and assuming normality of the ε_i term then,

$$y_{ji} = a_0 + a_1 P_{ij} + a_2 T_{ij} + e_{ij} \tag{1.6}$$

Rearranging Eqn. 1.6 to evaluate the subjective value of time,

$$y_{ij} = a_0 + a_1\left(P_{ij} + \frac{a_2}{a_1} T_{ij}\right) + e_{ij} \tag{1.7}$$

The subjective value of time or the willingness to pay is $q = a_2/a_1$. If the parameters a_1, a_2 are denoted as the estimates of the parameters a_1, a_2, then,

$$\bar{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \sim N(a, \Sigma)$$

where: $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$

$$\Sigma = \sigma^2 (X'X)^{-1}$$

$$X = \begin{pmatrix} 1 & P_1 & T_1 \\ \vdots & \vdots & \vdots \\ 1 & P_{ij} & T_{ij} \end{pmatrix}$$

2. The ratio of normal variables

The problem of the ratio of normal variables is common in the field of biomedical assay, bioequivalence, calibration and agriculture (for example, in the estimation of red cell life span and ratio of the weight of a component of the plant to that of the whole plant). The nature of the distribution of the ratio depends on the parameters, m_1 & m_2 (means), s_1 & s_2 (standard deviations) and r (correlation coefficient) of the bivariate normal distribution of the primary variables X and Y. Fieller (1932) and Hinkley (1969) derived the density function of the ratio as

$$f(\hat{q}) = \frac{hl}{\sqrt{2p}\, s_1 s_2 g^3} \left\{ 2\Phi\left(\frac{h}{g\sqrt{1-r^2}} \right) - 1 \right\} + \frac{\sqrt{1-r^2}}{p s_1 s_2 g^2} \exp\left\{ -\frac{k}{2(1-r^2)} \right\} \qquad (2.1)$$

where:

$$g = \left(\frac{\hat{q}^2}{s_2^2} - \frac{2r\hat{q}}{s_1 s_2} + \frac{1}{s_1^2} \right)^{1/2}$$

$$k = \frac{m_1^2}{s_1^2} - \frac{2r\hat{q} m_1 m_2}{s_1 s_2} + \frac{m_2^2}{s_2^2}$$

$$h = \left(\frac{m_2}{s_2} - \frac{r m_1}{s_1} \right) \left(\frac{\hat{q}}{s_2} \right) + \left(\frac{m_1}{s_2} - \frac{r m_2}{s_2} \right) \left(\frac{1}{s_1} \right)$$

$$l = \exp\left\{ \frac{1}{2(1-r^2)} \left(\frac{h^2}{g^2} - k \right) \right\}$$

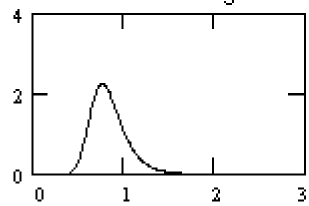
$$r = \frac{s_{12}}{s_1 s_2}$$

$$\Phi(\bullet) = \text{cdf } N(0,1)$$

The variations of the distribution depending on the values of means, variances (or standard deviation) and the correlation coefficient are discussed extensively in "Probability and Statistics" under the subheading "Ratio Populations" (available online: http://mathpages.com/home/kmath042/kmath042.htm). To illustrate, relevant sections and graphs are reproduced here.

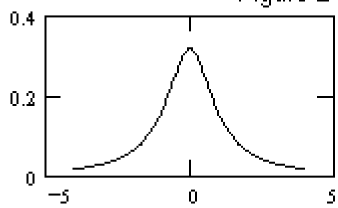
Suppose X is a normal population with mean of $\mu_1 = 90$ and standard deviation of $\sigma_1 = 12$, and suppose Y is a normal population with mean of $\mu_2 = 110$ and standard deviation of $\sigma_2 = 20$. A plot of the density distribution for this case is shown in Figure 1. This is not a normal distribution but the distribution is fairly well behaved to define a "pseudo mean" and "pseudo standard deviation".

Figure 1



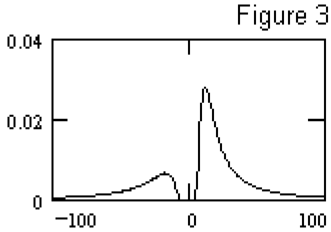
In the case of the ratio between two independently distributed standard normal variables, $X \sim N(0,1) / Y \sim N(0,1)$, it follows a Cauchy distribution. Cauchy distribution is unstable with indefinite variance and no mean. A plot of this is shown in Figure 2.

Figure 2



Marsaglia (1965) addressed the fundamental formulation of the distribution of the ratio of normal variables, as well as pointed out the potential for the ratio probability density function to exhibit bimodal behavior. Now for an interesting example, suppose the X population has a mean of 20 and a standard deviation of 1 and suppose the Y population has a mean of 0.5 and a standard deviation of 1. The density of the ratio is shown in Figure 3.

Figure 3



To summarize, the distribution of ratios of two normal distributions could either be skewed unimodal (single peak) or bimodal (two peaks) depending on the values of m_1 & m_2 (means), s_1 & s_2 (standard deviations) and r (correlation coefficient).

3. Methods of building the confidence interval

The current methods of building the confidence interval applied in value of time studies do not use directly the probability density function of the ratio. Refer to Armstrong, Garrido and Ortúzar (2001) for further discussion of these methods.

This study highlights the direct substitution method wherein the probability density function of the ratio is utilized and the estimation by t-test method.

(1) Direct substitution method

Estimated values of μ, σ, and ρ from actual data are plugged into Eqn. 2.1.

$$E(a_1) = m_1 \qquad V(a_1) = s_1^2 \qquad r = \frac{s_{12}}{s_1 s_2}$$
$$E(a_2) = m_2 \qquad V(a_2) = s_2^2$$

From the resulting graph of the distribution, the confidence interval is computed given a 95% confidence limit,

$$\Pr[q^2 < q_{0.95}^2] = 0.95 \qquad (3.1)$$

such that,

$$q_{0.025} < \hat{q} < q_{0.975} \qquad (3.2)$$

(2) t-test method

The t-test method is an instance of what statisticians call the *method of pivots*, wherein a pivot is a function of the data and the parameters whose distribution is independent of the value of the true parameter. The more prominent example of the method of pivots is the Fieller's method. The t-test method is slightly sophisticated since it uses Student t-distribution, to account for population variances (which need to be estimated by sample variances).

Suppose the linear statistics a_1 and a_2 are jointly normally distributed with expectations $E[a_1]=a_1$ and $E[a_2]=a_2$. Then,

$$q = E\left(\frac{a_2}{a_1}\right) = \frac{a_2}{a_1} \qquad (3.3)$$

Whatever the true value of θ,

$$Z = a_2 - qa_1 \sim N(0, s_z^2) \qquad (3.4)$$

where:

$$s_z^2 = s_2^2 - 2q_{12} + q^2 s_1^2$$
$$V(a_1) : s_1^2$$
$$V(a_2) : s_2^2$$
$$\operatorname{cov} a_1 a_2 = s_{12}$$

Let s_z^2 as the unbiased estimator of s_z^2, then the t-statistic is

$$t = \frac{a_2 - qa_1}{s_z} \qquad (3.5)$$

and

$$s_z^2 = s_2^2 - 2q_{12} + q^2 s_1^2 \qquad (3.6)$$

Setting the confidence level as 95%,

$$\Pr(t^2 < t_0^2) = 0.95 \qquad (3.7)$$

From Eqn.3.5 and given the condition $t^2 < t_0^2$,

$$\Pr\left[\frac{(a_2 - qa_1)^2}{(s_2^2 - 2qs_{12} + q^2 s_1^2)} < t_0^2\right] = 0.95 \qquad (3.8)$$

$$(a_1^2 - s_1 t_0^2 - 2q(a_1 a_2 - s_{12} t_0^2) + 2qs_{12} < 0 \qquad (3.9)$$

Solving for q,

$$\frac{a_1 a_2 - s_{12} t_0^2 - \sqrt{D/4}}{a_1^2 - s_1^2 t_0^2} < q < \frac{a_1 a_2 - s_{12} t_0^2 + \sqrt{D/4}}{a_1^2 - s_1^2 t_0^2} \qquad (3.10)$$

where:

$$D/4 = \left(s_{12}^2 - s_1^2 s_2^2\right) t_0^4 + \left(a_1^2 s_2^2 - 2a_1 a_2 s_{12} + a_2^2 s_1^2\right) t_0^2 \qquad (3.11)$$

4. Application to value of time

The data is composed of samples of commuter mode choice and the resulting estimates for the values of μ and σ are the following:

$$m_1 = -5.50x10^{-5} \qquad s_1^2 = 3.49x10^{-12}$$
$$m_2 = -1.04x10^{-2} \qquad s_2^2 = 9.91x10^{-9}$$
$$s_{12} = -6.50x10^{-11}$$

The distribution of the ratio is plotted by solving the direct substitution method and is shown in Fig. 4. The graph is a skewed unimodal case (single peak).

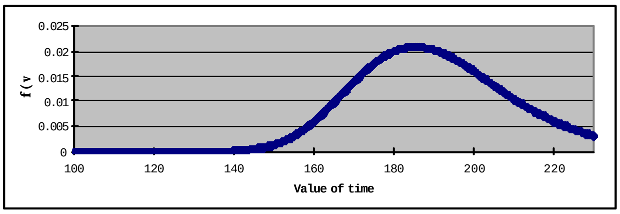


Fig.4 The distribution of the ratio of value of time

The confidence intervals (lower bound and upper bound), computed by the direct substitution method and the estimation by t-test method, are shown in Table 1.

(1) Comparative statics

Comparative statics was also done by varying the values of σ_1 and σ_2 while the covariance value σ_{12} is fixed as $-6.50x10^{11}$ as shown in Table 1. The left side values are the lower bound (L) while the right side values are the upper bound (U) of the confidence interval.

Table 1 Comparison of direct substitution method and t-test

	σ_2	X1.5		X2		X3	
σ_1		L	U	L	U	L	U
X 1.5	direct	165.3	208.4	164.6	208.8	162.7	210.1
	t-test	170.5	211.7	169.9	212.2	168.5	213.7
X2	direct	159.5	214.8	159.0	215.2	157.4	216.2
	t-test	165.8	219.4	165.4	219.8	164.2	221.0
X3	direct	148.8	229.3	148.5	229.5	147.4	230.2
	t-test	156.9	237.2	156.6	237.4	155.8	238.3

From Table 1, the confidence interval derived from the direct substitution method is lower and narrower as compared to the results from the estimation by the t-test method.

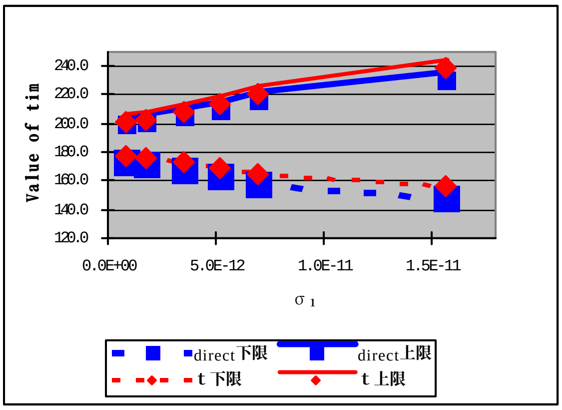


Fig.5 Confidence interval with varying values of σ_1

Further, Fig. 5 shows the divergence of the values with the varying of the values of σ_1. Almost the same figure is attained by varying σ_2.

(2) Minimization of width of the interval (t-test method)

The theory does not prescribe exactly how to choose the endpoints for the confidence interval. An obvious criterion is to minimize the width of the interval (Greene, 2003). To minimize the width of the confidence interval from the t-test method, set

$$\Pr(|X| < t_a) = 2a \quad (4.1)$$

$$\Pr(|X| < t_{1-a}) = 2(5-a) \quad (4.2)$$

Then, the resulting value of q is solved by

$$\frac{a_1 a_2 - s_{12} t_0^2 - \sqrt{D_1/4}}{a_1^2 - s_1^2 t_0^2} < q < \frac{a_1 a_2 - s_{12} t_0^2 + \sqrt{D_2/4}}{a_1^2 - s_1^2 t_0^2} \quad (4.3)$$

where:

$$D_1/4 = \left(s_{12}^2 - s_1^2 s_2^2\right) t_1^4 + \left(a_1^2 s_2^2 - 2a_1 a_2 s_{12} + a_2^2 s_1^2\right) t_1^2 \quad (4.4)$$

$$D_2/4 = \left(s_{12}^2 - s_1^2 s_2^2\right) t_2^4 + \left(a_1^2 s_2^2 - 2a_1 a_2 s_{12} + a_2^2 s_1^2\right) t_2^2 \quad (4.5)$$

After minimizing the width of the confidence interval of the t-test method, the difference between the results from the direct substitution method was eliminated. Refer to Table 2 for the comparison between the direct substitution method and the t-test method with minimized width of interval.

Table 2 Comparison of direct substitution method and t-test with minimized interval width

	σ_2	X1.5		X2		X3	
σ_1		L	U	L	U	L	U
X 1.5	direct	169.5	210.5	168.9	211.0	167.3	212.3
	t-test	169.5	210.5	169.0	211.1	167.6	212.5
X2	direct	164.0	217.2	164.0	218.0	162.5	218.8
	t-test	164.2	219.3	163.8	217.8	162.6	218.9
X3	direct	153.4	232.0	153.2	232.4	152.5	233.3
	t-test	153.7	232.3	153.4	232.6	152.6	233.4

5. Conclusions

The probability density function (pdf) of the ratio of normal variables is a well-solved problem and could be used by direct substitution of the estimated values of μ, σ and ρ to derive the distribution of the ratio. However, considering the complexity of the pdf of the ratio, it is worthwhile to consider estimation of the confidence interval with a certain probability level. The applicability of the t-test method in building the confidence interval to the value of time is discussed and shown in the study. For further research is the suitability of the t-test method to the bimodal case.

Based on the comparison of the results from the two methods, the direct substitution method yields a lower and narrower confidence interval. However, by applying minimization of the width of the confidence interval to the t-test method, the discrepancy was minimized if not eliminated.

References

1) Armstrong, P., Garrido, R., Ortúzar, J.: Confidence interval to bound the value of time, Transportation Research Part E, Vol.37, pp.143-161, 2001.

2) Ben-Akiva and Lerman: Discrete choice analysis: theory and application to travel demand, MIT Press, Cambridge, MA, 1985.

3) Fieller, E.: The distribution of the index in a normal bivariate population, Biometrika, XXIV, pp. 428-440, 1932.

4) Garrido, R. and Ortúzar, J.: The Chilean value of time study: methodological developments, Proceedings of the 21st PTRC Summer Annual Meeting, University of Manchester Institute of Science and Technology, England, 1993.

5) Greene, W.H.: Econometric analysis, Prentice Hall, Upper Saddle River, NJ, 2003.

6) Heitjan, D.F.: Fieller's method and net health benefits, Health Economics, Vol.9, pp.327-335, 2000.

7) Hinckley, D.V.: On the ratio of two correlated normal random variables, Biometrika, Vol.56, pp.635-639,1969.

8) Marsaglia, G.: Ratios of normal variables and ratios of sums of uniform variables, Journal of the American Statistical Association, Vol.60, pp.193-204, 1965.