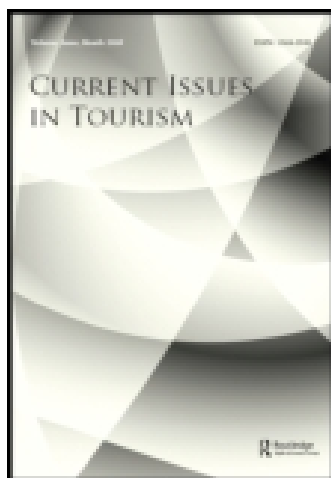


This article was downloaded by: [National Chiao Tung University 國立交通大學]

On: 24 December 2014, At: 17:32

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Current Issues in Tourism

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rcit20>

An accurate confidence interval for the mean tourist expenditure under stratified random sampling

Cherng G. Ding^a & Hsiu-Yu Lee^a

^a Institute of Business and Management, National Chiao Tung University, 118 Chung-Hsiao West Road, Section 1, Taipei 100, Taiwan, Republic of China

Published online: 12 Nov 2013.



CrossMark

[Click for updates](#)

To cite this article: Cherng G. Ding & Hsiu-Yu Lee (2014) An accurate confidence interval for the mean tourist expenditure under stratified random sampling, *Current Issues in Tourism*, 17:8, 674-678, DOI: [10.1080/13683500.2013.857296](https://doi.org/10.1080/13683500.2013.857296)

To link to this article: <http://dx.doi.org/10.1080/13683500.2013.857296>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

CURRENT ISSUES IN TOURISM LETTER

An accurate confidence interval for the mean tourist expenditure under stratified random sampling

Cherng G. Ding* and Hsiu-Yu Lee†

Institute of Business and Management, National Chiao Tung University, 118 Chung-Hsiao West Road, Section 1, Taipei 100, Taiwan, Republic of China

(Received 4 September 2013; final version received 15 October 2013)

It is of interest to make an inference about the mean expenditure per tourist per day (or per stay). A simple approach to obtain an exact t test under stratified random sampling is proposed in this study. The underlying assumptions include given prior probabilities, normality within each stratum, and the equality of within-stratum variances. Procedure to make an exact inference has been provided and illustrated. The proposed approach is effective and recommended for the studies where stratified random sampling is conducted and inference for the mean expenditure is desirable.

Keywords: tourist expenditure; stratified random sampling; exact t test

Introduction

With increasing globalisation, tourism plays a significant role in economic development. Over the years, tourism and its associated economic repercussions have taken place within a wider context of globalisation of the world economy (Sugiyarto, Blake, & Sinclair, 2003). For many countries, areas, and cities, tourism has become an important source for generating revenues, employment, infrastructure development, and economic growth (Lee & Chang, 2008), and tourist expenditure has been a key variable for evaluating tourism's economic effects (Aguilo Perez & Juaneda, 2000; Dwyer, Forsyth, & Spurr, 2004).

When a tourism project is proposed, it is desirable to have information about both risk and benefits, which can be assessed based on the expected expenditure per tourist per day (or per stay), regardless of domestic tourists, international tourists, or both. Providing a range to reflect precision of an estimate of the expected tourist expenditure is more informative than a single estimate. This can be achieved by applying the statistical interval estimation rather than point estimation. Approximate confidence intervals for tourist expenditure by using bootstrapping, a nonparametric resampling procedure, have been available (English, 2000; Pol, Pascual, & Vazquez, 2006). A frequently used sampling technique for tourism survey is the stratified random sampling, a more efficient sampling technique than the simple random sampling when data are more homogeneous within each stratum than in the entire population (Cochran, 1977). Within each stratum determined based on tourist background (e.g. residence, travelling purpose), tourist expenditure data are collected from randomly selected tourists. However, how to obtain an accurate interval

*Corresponding author. Email: cding@mail.nctu.edu.tw

†Current address: Accounting and Statistics Division, Ministry of Culture, Republic of China, 30-1 Beijing East Road, Taipei 100, Taiwan, Republic of China.

estimate for the mean tourist expenditure under stratified random sampling has received little attention in tourism-related research. The purpose of this note is to fill this gap by proposing an approach to make an exact inference, which can improve the quality of subsequent analysis, lead to more accurate investment assessment and economic analysis, and in turn result in more efficient resource allocation.

An exact *t* test

Let a population consist of L strata and p_i denote the prior probability for stratum i for which $\sum_{i=1}^L p_i = 1$. Under stratified random sampling, L simple random samples are obtained independently. A stratified random sample contains L independent subsamples from different strata. Each subsample is a simple random sample, composed of random variables which are independent and identically distributed. Let the distribution of tourist expenditure within stratum i be $N(\mu_i, \sigma_i^2)$. Let n_i, \bar{X}_i and S_i^2 denote, respectively, the sample size, the sample mean and the sample variance for stratum i . It is well known that \bar{X}_i is $N(\mu_i, \sigma_i^2/n_i)$, $(n_i - 1) S_i^2/\sigma_i^2$ is distributed as χ^2 with $(n_i - 1)$ degrees of freedom, and \bar{X}_i and S_i^2 are independent. The population mean (the mean for the entire population) is given by $\mu = \sum_{i=1}^L p_i \mu_i$. To estimate μ , use the stratified mean $\bar{X}_{st} = \sum_{i=1}^L p_i \bar{X}_i$ (Cochran, 1977, p. 91). \bar{X}_{st} is distributed as $N(\mu, \sum_{i=1}^L p_i^2 \sigma_i^2/n_i)$. To test $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$, use $t_{appr} = (\bar{X}_{st} - \mu_0)/\sqrt{\sum_{i=1}^L p_i^2 S_i^2/n_i}$, which is approximately distributed as t with f degrees of freedom, where $f = (\sum_{i=1}^L p_i^2 S_i^2/n_i)^2 / (\sum_{i=1}^L p_i^4 S_i^4 / (n_i^2 (n_i - 1)))$ (Cochran, 1977, p. 96; Govindarajulu, 1999, pp. 78–79). The associated $100(1 - \alpha)\%$ confidence interval for μ is given by $[\bar{X}_{st} - t_{\alpha/2;f} \sqrt{\sum_{i=1}^L p_i^2 S_i^2/n_i}, \bar{X}_{st} + t_{\alpha/2;f} \sqrt{\sum_{i=1}^L p_i^2 S_i^2/n_i}]$, where $t_{\alpha/2;f}$ denotes the upper $\alpha/2$ percentage point of the t distribution with f degrees of freedom.

If it is further assumed that the within-stratum variances are equal, then \bar{X}_{st} is $N(\mu, \sum_{i=1}^L p_i^2 \sigma^2/n_i)$, and $(\bar{X}_{st} - \mu) / (\sqrt{\sum_{i=1}^L p_i^2/n_i} \sigma)$ is $N(0, 1)$. In addition, $\sum_{i=1}^L (n_i - 1) S_i^2/\sigma^2$ is distributed as χ^2 with $(n - L)$ degrees of freedom, where $n = \sum_{i=1}^L n_i$, and is independent of \bar{X}_{st} . It follows that (see also Kutner, Nachtsheim, Neter, & Li, 2005, Section 17.3) $(\bar{X}_{st} - \mu) / (\sqrt{\sum_{i=1}^L p_i^2/n_i} \sqrt{\sum_{i=1}^L (n_i - 1) S_i^2 / (n - L)})$ is distributed as t with $n - L$ degrees of freedom. In fact, $\sum_{i=1}^L (n_i - 1) S_i^2 / (n - L)$ is a pooled estimator, denoted by S_p^2 , for the common within-stratum variance σ^2 , and $\sqrt{\sum_{i=1}^L p_i^2/n_i} S_p$ is the estimated standard error of \bar{X}_{st} . To test $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$, compute $t_{exact} = (\bar{X}_{st} - \mu_0) / (\sqrt{\sum_{i=1}^L p_i^2/n_i} S_p)$, and reject H_0 at the α significance level if $|t_{exact}| \geq t_{\alpha/2; n-L}$, where $t_{\alpha/2; n-L}$ denotes the upper $\alpha/2$ percentage point of the t distribution with $n - L$ degrees of freedom; otherwise H_0 is not rejected. The associated $100(1 - \alpha)\%$ confidence interval for μ is given by $[\bar{X}_{st} - t_{\alpha/2; n-L} \sqrt{\sum_{i=1}^L p_i^2/n_i} S_p, \bar{X}_{st} + t_{\alpha/2; n-L} \sqrt{\sum_{i=1}^L p_i^2/n_i} S_p]$. For the one-sided alternative hypothesis $H_1: \mu > \mu_0$ (or $\mu < \mu_0$), reject H_0 if $t_{exact} \geq t_{\alpha; n-L}$ (or $t_{exact} \leq -t_{\alpha; n-L}$).

With proportional allocation, p_i is equal to n_i/n , $\sum_{i=1}^L p_i^2/n_i$ becomes $1/n$, and t_{exact} reduces to $t_{\text{exact}} = (\bar{X}_{\text{st}} - \mu_0)/(S_p/\sqrt{n})$. The associated $100(1-\alpha)\%$ confidence interval for μ is given by $[\bar{X}_{\text{st}} - t_{\alpha/2;n-L}(S_p/\sqrt{n}), \bar{X}_{\text{st}} + t_{\alpha/2;n-L}(S_p/\sqrt{n})]$.

To compute the power of the test based on t_{exact} when H_0 is not true, we need the distribution of t_{exact} under H_1 . When the population distribution is normal and variances within strata are equal, the theoretical exact power of the test under H_1 , i.e. $P(|t_{\text{exact}}| \geq t_{\alpha/2;n-L}|H_1)$ is evaluated based on the noncentral t distribution with $n - L$ degrees of freedom and non-centrality parameter $\theta = (\mu^* - \mu_0)/\left(\sqrt{\sum_{i=1}^L p_i^2/n_i} \sigma\right)$, where $\mu^* \in H_1$ (Larsen & Marx, 1986, Appendix 7.1). Note that t_{exact} and θ are independent of the between-stratum variation.

The assumptions underlying the exact t test include given prior probabilities, the normality of tourist expenditure within each stratum and the equality of within-stratum variances of tourist expenditure. The assumptions must be satisfied to validate the test. The t_{exact} test is more powerful than the t_{appr} test. The sample size needed to achieve a specified power is smaller for the former. The confidence interval for μ based on t_{exact} is more accurate than that based on t_{appr} . In addition, t_{exact} is simpler than t_{appr} since its degrees of freedom do not depend on within-stratum variances, and hence it is easier to deal with the sample size problems based on t_{exact} .

Under stratified random sampling, a simple procedure to implement the approach, assuming that prior probabilities are given, is as follows:

Step 1: Test for the normality of tourist expenditure within each stratum.

Normality can be tested by using the tests available in statistical software such as the Shapiro–Wilk W test (Shapiro & Wilk, 1965). However, practitioners need to be aware that the tests do not perform well for small sample sizes (≤ 30) (Razali & Wah, 2011).

Step 2: Test for the equality of within-stratum variances of tourist expenditure.

Equality of variances can be examined by the Brown–Forsythe test (Kutner et al., 2005, p. 784).

Step 3: If the conditions specified in the previous two steps both hold, then make inference for the population mean of tourist expenditure by using the exact t statistic.

Illustration

The Taiwan Tourism Bureau has been promoting tourism projects to attract international tourists. In this respect, both public and private tourism sectors require meaningful and accurate estimates of international tourism demand in order to fulfil efficient allocation of limited resources. Since accurate inference for tourist expenditure is essential for further tourism investment analysis and planning, the proposed approach should be useful.

To illustrate the inferential procedure, we use the 2011 inbound tourist expenditure survey data collected by the Taiwan Tourism Bureau. The data set includes the variable of the travelling purpose (consisting of recreation, business, and others that do qualify as tourism (Weaver & Oppermann, 2000, p. 29)). Since it is believed that the mean daily expenditure differs among travelling purposes, the travelling purpose can be used as the stratification criterion. Suppose it is desired to obtain the mean daily expenditure of Japanese tourists for March, 2011. The information about the inference is summarised in Table 1. There exist three strata: recreation, business and others, with the prior

Table 1. Information for the estimation of the mean Japanese tourist daily expenditure in Taiwan for March, 2011.

	Stratum (travelling purpose)		
	Recreation	Business	Others
Prior probability (p_i)	0.2140	0.7015	0.0845
Sample size (n_i)	54	13	9
Sample mean (\bar{X}_i)	341.47	329.63	208.44
Sample variance (S_i^2)	32326.04	45112.59	40301.20
Test for normality	0.0600 ^a	0.4685 ^a	0.1535 ^a
Test of the equality of within-stratum variances: 0.7862 ^b			
95% confidence interval for μ based on t_{exact} : [\$247.51 USD, \$396.34 USD]			

^aThe p -value of the Shapiro–Wilk W test.

^bThe p -value of the Brown–Forsythe test.

probabilities of 0.2140, 0.7015, and 0.0845, respectively, obtained from the population data provided by the Taiwan Tourism Bureau. The procedure to make an inference is given as follows:

Step 1: The p values associated with the Shapiro–Wilk W test for the three travelling purposes were, respectively, 0.0600, 0.4685, and 0.1535, supporting the normality of tourist daily expenditure within each stratum.

Step 2: The p value associated with the Brown–Forsythe modified test was 0.7862, supporting the equality of within-stratum variances.

Step 3: Since the assumptions of normality within each stratum and equality of within-stratum variances have been satisfied, we can proceed to compute the confidence interval for the mean daily expenditure per tourist based on t_{exact} . The exact 95% confidence interval for μ is given by [\$247.51 USD, \$396.34 USD].

Confidence intervals for the mean tourist daily expenditure for different populations at different time periods can be obtained in a similar way. The information obtained is useful for tourism policy-makers, the tourism industry, and tourists.

Conclusion

An approach for constructing an exact confidence interval for the mean tourist expenditure has been provided and illustrated with real expenditure data. The assumptions underlying the approach include given prior probabilities, the normality within each stratum, and the equality of within-stratum variances. The approach is effective and easy to implement. It is recommended for tourism studies. Robustness of the approach needs to be further studied.

Acknowledgements

The authors thank the Taiwan Tourism Bureau for providing the survey data and Professors Hubert J. Chen and Gwonen Shieh for helpful discussions. The authors also thank an anonymous reviewer for valuable suggestions.

References

- Aguilo Perez, E., & Juaneda, S. C. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research*, 27, 624–637.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley & Sons.
- Dwyer, L., Forsyth, P., & Spurr, R. (2004). Evaluating tourism's economic effects: New and old approaches. *Tourism Management*, 25, 307–317.
- English, D. B. K. (2000). A simple procedure for generating confidence intervals in tourist spending profiles and resulting economic impacts. *Journal of Regional Analysis & Policy*, 30, 59–74.
- Govindarajulu, Z. (1999). *Elements of sampling theory and methods*. Upper Saddle River, NJ: Prentice-Hall.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). Singapore: McGraw-Hill International Edition.
- Larsen, R. J., & Marx, M. L. (1986). *An Introduction to mathematical statistics and its applications* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Lee, C. C., & Chang, C. P. (2008). Tourism development and economic growth: A closer look at panels. *Tourism Management*, 29, 180–192.
- Pol, A. P., Pascual, M. B., & Vazquez, P. C. (2006). Robust estimators and bootstrap confidence intervals applied to tourism spending. *Tourism Management*, 27, 42–50.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *Journal of Statistical Modeling and Analytics*, 2, 21–33.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.
- Sugiyarto, G., Blake, A., & Sinclair, M. T. (2003). Tourism and globalization economic impact in Indonesia. *Annals of Tourism Research*, 30, 683–701.
- Weaver, D., & Oppermann, M. (2000). *Tourism management*. New York, NY: John Wiley & Sons.