HHS Public Access

Median and Quantile Tests under Complex Survey Design using SAS and R

Yi Pan[1], **Sam Caudill**[1], **Ruosha Li**[2] **[Assistant professor]**, and **Kathy L. Caldwell**[1]

[1]Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia.

[2]Department of Biostatistics, University of Pittsburg, Pittsburgh, PA 15261.

Abstract

The programs on testing the medians or other quantiles of two or more subgroups under complex survey design are limited. In this paper, we introduce programs in both SAS and R to perform such a median/quantile test under complex survey design. A detailed illustration of the computations, macro variable definitions, input and output for the SAS and R programs are also included in the text. Urinary iodine data from National Health and Nutrition Examination Survey (NHANES) are used as examples for comparing medians between females and males as well as comparing the 75th percentiles among three salt consumption groups.

Keywords

Median; Quantile; Complex Survey; SAS; R

1. Introduction

National complex survey results are the basis of many public health related policies, these results are usually from a complex finite population sample,, possibly stratified clustered and unequally weighted. Therefore, analyses of health survey data obtained with complex sampling designs are extraordinarily important. Natarjan et al. did a search on PubMed (National Library of Medicine) and they found 7699 articles related to "NHANES" (National Health and Nutrition Examination Survey) between the years 2006 and 2011 [1]. However, inference methods and available software for testing medians or quantiles of 2 or more groups are not well developed for complex survey data. In this paper, we address this issue and introduce a SAS Macro and an R function to perform the quantile tests under complex survey design.

Several statistical approaches have been proposed to test for the equality of medians or different quantiles, but these tests do not take into account prior the extension to the complex survey design. Brown and Mood ([2–3]), Mood [4–5] and Westenberg [6] introduced Mood's median test, which is also referred as the joint median test. The median test is a distribution-free procedure and focuses on whether two or more independent samples differ in their median values for a criterion variable. And the median test is a special case of a joint quantile test [7]. In the meantime, the Wilcoxon rank-sum test (also called the Wilcoxon-

Mann-Whitney test) was proposed as an alternative [8–10]. Few studies extend those different median/rank tests under complex sampling. Williams and Perritt [11] investigated a simple way to extend Mood's median test to a weighted median test under a complex survey design, but it was restricted to only compare medians. Natarajan et al. [1] discussed the relation between the proportional odds model and the Wilcoxon rank sum test. By formulating the Wilcoxon test statistics in terms of a score test statistic from the proportional odds model, they extended the Wilcoxon rank-sum test to handle complex samples, obtaining the test statistic as the score test statistic for the group effect from a proportional odds cumulative logistic regression model. Furthermore, Lumley and Scott [12] investigated a general approach to construct design-based rank tests when comparing two groups within a complex sample. Their approach is incorporated in the R survey package "survey" [13] but it only applies to testing medians and quantiles between two groups.

In summary, the computer programs for testing the equality of different quantiles for two or more groups are limited. In this paper, we present a quantile test based on the extension of mood's median test for two or more groups under a clustered and stratified (complex) survey design by both SAS macro and an R function. The programs will benefit researchers who are working with medians or other quantiles from survey data. Computational methods and theory are introduced in the next section. Section 3 describes the details of the SAS and R programs.. Two examples from NHANES are included to illustrate the quantile test on two or more groups. A brief summary follows in section 5.

2. Methods

2.1 Mood's Median Test

Mood's median test compares the medians of observations from two or more samples and is very straightforward to follow. The test assumes that observations are independent both within and between samples, and the distributions of the populations the samples were drawn from all have the same shape.

Following the notation of Williams and Perritt [11], we consider a population of N observations with C samples, N_1, N_2, …, N_C. Define M as the median of all N observations and M_i as the median of the ith sample. In median test, we're interested in the following hypotheses.

$$H_0 : M_1 = M_2 = \cdots = M_c = M$$

$$H_1 : \text{at least } M_i \neq M_j, \text{where } i \neq j.$$

Furthermore, define A_{i1} as the number of observations in the ith sample less than or equal to M. On the other hand, let A_{i2} denote the number of observations in the ith sample greater than M.

	N_1	N_2	...	N_c	N
M	A_{11}	A_{12}	...	A_{1c}	K
$> M$	A_{21}	A_{22}	...	A_{2c}	K

And let $A_{11}+A_{12}+\cdots+A_{1c}=A_{21}+A_{22}+\cdots+A_{2c}=K=\frac{1}{2}N$.

If the null hypothesis is true, any given observation will have a probability 0.5 of being greater than the overall median M, by definition and regardless of which sample it is from. Therefore, for each sample, the number of observations greater than M will have a binomial distribution with $p = 0.5$.

Define P_i as the probability that a certain observation in the ith sample is less than or equal to M. Finally, the median test hypotheses can be transformed as follows:

$$H_0:P_1=P_2=\cdots=P_c=\frac{1}{2}$$

$$H_1:\text{at least } P_i \neq P_j \neq \frac{1}{2}, \text{where } i \neq j.$$

The above hypotheses can be tested using a large sample chi-square statistic with $C-1$ degrees of freedom.

2.2 Extension to Quantile Test and Complex Survey Samples

The median test is a special case of a quantile test [7]. If we define Q as the value of qth quantile on the overall population. The contingency table in Section 2.1 can be modified as follows.

	N_1	N_2	...	N_c	N
Q	A_{11}	A_{12}	...	A_{1c}	K_1
$> Q$	A_{21}	A_{22}	...	A_{2c}	K_2

The hypotheses in a quantile test are:

$$H_0:P_1=P_2=\cdots=P_c=q, \text{where } q \in (0, 1),$$

$$H_1:\text{at least } P_i \neq P_j \neq q, \text{where } i \neq j.$$

However, extending this framework to a complex survey design remains challenging. A complex sample is a probability sample based on sampling procedures such as stratification, clustering and weighting, and designed to improve statistical efficiency, reduce costs or improve precision for subgroup analyses relative to simple random sampling. And independence of observations and equal probabilities of selection may no longer hold under a complex sampling design.

Let us consider a stratified, clustered complex sampling design and define $y_{h\alpha i}$ as the measurements on unit i in cluster α and in stratum h. Also following the notation from Williams and Perritt [11], given the distribution F of $y_{h\alpha i}$ in the entire population, Q is the value of the qth population quantile such that $F(Q-) < q \leq F(Q)$. Define the sample empirical distribution function as $G(y) = \dfrac{\sum_h \sum_\alpha \sum_i w_{h\alpha i} I_y(y_{h\alpha i})}{\sum_h \sum_\alpha \sum_i w_{h\alpha i}}$, where $w_{h\alpha i}$ is the corresponding weight and I_y is the indicator function such that $I_y(y_{h\alpha i}) = \begin{cases} 1 \text{ if } y_{h\alpha i} \leq y \\ 0 \text{ if } y_{h\alpha i} > y \end{cases}$.

The qth population quantile can be consistently estimated by Q, which satisfies $G(Q-)<q \leq G(Q)$

To perform the quantile test based on the extended Mood's median test, we want to estimate $P_{h\alpha i}$ as follows: $p_{h\alpha i} = \dfrac{\sum_h \sum_\alpha \sum_i w_{h\alpha i} x_{h\alpha i|c} I_Q(y_{h\alpha i})}{\sum_h \sum_\alpha \sum_i w_{h\alpha i} x_{h\alpha i|k}}$, where $I_Q(y_{h\alpha i})$ is the indicator function we defined above, with y replaced by Q, and $x_{h\alpha i|c}$ is the indicator subgroup c for unit i in cluster α and in stratum h. $x_{h\alpha i|c} = \begin{cases} 1 \text{ if unit is in subgroup } c \\ \quad 0 \text{ otherwise.} \end{cases}$

In order to obtain a consistent estimate of the covariance matrix V of $p = [p_1, p_2, p_3, \ldots, p_c]'$ under a complex survey design, we regress $I_Q(y_{h\alpha i})$ onto $x_{h\alpha i|c}$, using a model that takes the survey design into account. Linear regression and logistic regression give identical point estimates. And based on limited simulations (results not shown) with various quantiles (i.e., 0.50, 0.75, 0.90, 0.95, and 0.99), the variance and covariance estimates from linear regression models differ from those obtained from logistic regression models by less than 0.2%, essentially resulting in no hypothesis testing differences between the two models. So for computational simplicity, we choose to use a linear regression model as suggested by Williams and Perritt [11]. Therefore, the general hypotheses can be presented as follows:

$$H_0 : P_1 = P_2 = \cdots = P_c = q$$

$$H_1 : \text{at least } P_i \neq P_j \neq q, \text{where } i \neq j.$$

The test statistic for such a quantile test is given by $T = [CP]'[CVC']^{-1}[CP]$, where $C = [J_{C-1}| - I_{C-1}]$, J_{C-1} is a column vector of all ones and I_{C-1} is the identity matrix with dimension $C - 1$. T is approximately distributed as a chi-squared distribution with $C - 1$ degree of freedom under H_0.

3. SAS and R Programs

Both SAS macro and R program can be used to test the equality of any quantile for 2 or more subgroups, under a clustered, stratified and weighted survey design. In this section, we introduce the input and output information for each program. Before using either program, it is required to calculate the quantile of the variable of interest over the entire population, using popular statistical software such as SUDAAN, SAS, R or STATA.

3.1 SAS macro

In the SAS macro, eight parameters, namely *data, analyte, variable, group, quantile, weight, strata* and *cluster* are required for input. The variable data specifies the dataset being analyzed and it is required to have the sampling weight of the variable of interest, strata, cluster, analyte and variable. The macro variable "analyte" is the name of the continuous variable of interest and "variable" is the name of the categorical variable which contains subgroup information; it should range from 1, 2, …, group depending on the number of categories. The variable "group" is the total number of groups in the categorical variable. For example, when comparing the urinary iodine levels between females and males, "group"=2. Furthermore, "quantile" is the pre-calculated quantile value of interest; "weight" is the sampling weight assigned to each sample person. It is a measure of the number of people in the population represented by that sample person. The variable "strata" defines homogenous groups. Dividing the population into relatively homogenous groups (strata) and sampling a predetermined number from each stratum will increase precision for a given sample size [14]. The variable "cluster" names the variables that identify the clusters in a complex survey design. Dividing the population into groups and sampling from a random subset of these groups (e.g. geographical locations) will decrease precision for a given sample size but often increase precision for a given cost [14].

The following output is provided by the SAS macro: (1) P; (2) SE_P; (3) COVV_P and; (4) Chi and (5) P Value. Based on the notation in Section 2, testing the equality of quantiles is comparable to testing for equal proportions. P is a vector with (C−1) dimensions. "SE_P" and "COVV_P" are the variance and covariance structure of the P values. "Chi" denotes the test statistic calculated based on T in Section 2 and the corresponding "P Value" shows the significance of the test.

3.2 R Program

In R function "test.quan", we specify 3 variables, namely *dataset, group* and *quantile*. The variable "dataset" is the dataset which contains "variable", "cluster", "strata", "analyte" and "weight" in the exact order we listed. The column names in "dataset" are also case sensitive. Furthermore, "group" is the total level of categorical variables. Finally, the variable "quantile" is defined the same as in the SAS macro, which is a pre-calculated quantile value relative to the entire population of interest. Note that any observations with missing weight should be removed from the dataset and any missing value on any variables should be presented as blank. The output from R will exactly match that from SAS.

4. Examples

In this section, we demonstrate two examples with both SAS and R programs input and output in Tables 1 and 2. Both examples are from NHANES 2009–2010, which is a complex, multistage, area probability sample representative of the US non-institutionalized civilian population during 2009–2010 [14]. All respondents gave their informed consent, and the NHANES protocol was reviewed and approved by the National Center for Health Statistics (NCHS) Research Ethics Review Board. Data collection consisted of: 1) a screening visit; 2) an interview during which a wide battery of health related questions were asked; and 3) an examination consisting of direct standardized physical examinations, including body measurements and blood and urine collection, carried out in a mobile examination center (MEC). In the following examples, we're interested in urinary iodine (UI) measurements. Iodine is very essential for proper thyroid function and controls human biochemical reactions and metabolism. As UI can change dramatically within the same subject each day, the World Health Organization (WHO) defines nutritional iodine sufficiency for a population by median UI concentrations as follows: excessive iodine intake, >300 μg/L; more than adequate intake, 200–299 μg/L; adequate intake, 100–199 μg/L; mild iodine deficiency, 50–99 μg/L; moderate iodine deficiency, 20–49 μg/L; and severe iodine deficiency <20 μg/L [15]. Therefore, comparing the median UIs or certain percentiles of UI between or among possible subgroups is of interest.

4.1 Test the equality of Medians between two groups in NHANES, 2009–2010

We're very interested in assessing the equality of median UIs between females and males based on NHANES 2009–2010. Based on descriptive analysis, the median UI for females is 134.4ug/L (95% CI = (125.4–146.3)) and that for males is 152.2ug/L (138.6–171.1). Before performing the median test, we also need to calculate the overall median by pooling males and females together, which is 143.5ug/L. As shown in Table 1, once we specify the input command in both SAS macro and R function, we achieve the test result from both programs. We conclude that males experience significant higher median UI compared to females based on NHANES 2009–2010 (p=0.004).

4.2 Test the equality of 75th percentiles among three groups in NHANES, 2007–2008

In addition, we're interested in investigating whether salt use is associated with 75th percentile of UI measurements based on NHANES 2007–2008. In NHANES, there is one dietary question related to salt consumption: How often is ordinary salt or seasoned salt added in cooking or preparing foods in your household? Is it never, rarely, occasionally, or very often? And in the analysis, salt consumption is defined as "never or rarely", "occasionally" and "very often". And we would like to test the equality of 75th percentiles of UI by three salt consumption categories. The 75th percentile of UI for "never or rare" group is 291.6ug/L (95% CI=(249.0–325.7)), for "occasionally" category is 283.85ug/L (261.1–302.7) and for "very often" category is 274.05ug/L (259.1–302.7). Before performing the test, we calculated the overall median which is 281.7ug/L. The test indicates there is no significant difference in the 75th percentiles of UI based on salt consumption groups (p=0.5).

5. Summary

In this paper, we provide computer programs based on an extension of Mood's median test to test the equality of any quantile by subgroups under a complex survey design. We introduce both SAS macro and R function to perform such a test. Two examples demonstrate that our programs work well for various situations. It's known that Mood's median test has been criticized for low power and requires a large sample size to obtain the adequate power. We know that usually a complex survey will provide sufficient sample size so that inadequate power should not be an issue here. Note that the step of calculating the quantile is not included in our programs because it may vary depending on which program is used. For example, SUDAAN and R calculate quantiles differently. If we have a string of data such as 1, 1, 1, 2, 2, 2, 2, 2, 3 with equal sampling weights, SUDAAN gives the median as 1.30, considering the values as continuous and R gives 2, considering the values as discrete [17].

6. Macro Availability and Software Requirement

The programs in SAS and R for quantile tests under a complex survey design are available directly from the authors. The programs were written in SAS V9.3 and R 2.14.2. The R users need to download and install the packages "survey" from CRAN mirror.

Acknowledgments

References

1. Natarjan S, Lipsitz SR, Fitzmaurice GM, Sinha D, Ibrahim JG, Haas J, Gellad W. An extension of the Wilcoxon rank sum test for complex sample survey data. Journal of the Royal Statistical Society, Series C. 2012; 61:653–664.

2. Brown GW, Mood AM. Homogeneity of several samples. Amarican Statistician. 1948; 2:22.

3. Brown, GW., Mood, AM. Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press; 1951. On median tests for linear hypotheses; p. 159-166.

4. Mood, AM. Introduction to the Theory of Statistics. McGraw-Hill; 1950.

5. Mood AM. On the asymptotic efficiency of certain nonparametric two-sample tests. Ann. Math. Stat. 1954; 25:514–522.

6. Westenberg J. Significance test for median and interquartile range in samples from continuous populations of any form. Nederl. Akad. Wetensch., Proc. 1948; 51:253–261.

7. Chakraborti S, Laan p. Precedence tests and confidence bounds for complete data: an overview and some results. The Statistician. 1996; 45:351–369.

8. Kruskal WH. Historical Notes on the Wilcoxon Unpaired Two-Sample Test. Journal of the American Statistical Association. 1957; 52:356–360.

9. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin. 1945; 1:80–83.

10. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics. 1947; 18:50–60.

11. Williams, RL., Perritt, RL. Proceedings of the section on survey research methods. American Statistical Association; Washington, D.C.: 1986. The weighted median test for finite population samples; p. 676-678.

12. Lumley, T., Scott, A. Technique Report Department of Statistics. University of Auckland; 2012. Two-sample rank tests under complex sampling.

13. R package "Survey". [Accessed: 02/12/2013] 2012. http://cran.r-project.org/web/packages/survey/index.html

14. Lumley T. Analysis of complex survey samples. Journal of Statistical Software. 2004; 9:1–19.

15. About the National Health and Nutrition Examination Survey. U.S. Centers for Disease Control and Prevention. [Last accessed: 02/12/2012] 2012. http://www.cdc.gov/nchs/nhanes.htm

16. World Health Organization. Assessment of iodine deficiency disorders and monitoring their eliminationa guide for programme managers. 2008

17. Lumley, T. Complex Surveys: a guide to analysis using R. John Wiley & Sons; 2011.

Table 1

Test the Equality of Medians between Males and Females in NHANES, 2009–2010.

Description	Test the equality of median urinary iodine between females and males in NHANES 2009–2010. "variable" =1 is male while 2 is female. "analyte" denotes the urinary iodine measurements.
Original Dataset	variablecluster strata analyte weight 1 2 86 62.9 30020.66 2 2 85 147.1 223481.95 1 1 86 201.7 30628.62 1 1 75 319.7 29664.02
SAS code	libname nhanes "D:\data"; %include 'D:\Quatile Test\program\Weighted_quatile_test.sas'; %test_quan(dataset=nhanes.test1, analyte=analyte, variable=variable, group=2, quantile=143.5, weight=weight, strata=strata, cluster=cluster);
R code	data<- read.csv("D:/data/test1.csv", header=T) test.quan(dataset=data, group=2, quantile=143.5)
Output	P SE_P COVV_P Chi P_Value 0.469 0.0217 0.0005 0.0002 8.261 0.004 0.529 0.0174 0.0002 0.0003

Table 2

Test the Equality of 75th Percentiles among Three Groups in NHANES, 2007–2008

Description	Test the equality of 75th percentile of urinary iodine among three salt consumption groups (never or rare, occasionally and very often). "variable"=1 means never or rare consume salt; 2 means occasionally consume salt and 3 means very often consume salt. "analyte" still denotes urinary iodine.

Original Dataset	variable	cluster	strata	analyte	weight
	3	1	60	210.1	60045.77
	2	1	70	244.6	35353.21
	3	1	67	364.3	10074.15
	3	1	70	224.3	9234.06

SAS code	libname nhanes "D:\data";
	%include 'D:\Quatile Test\program\Weighted_quatile_test.sas';
	%test_quan(data=nhanes.test2, analyte=analyte, variable=variable, group=3, quantile=281.7, weight=weight, strata=strata, cluster=cluster);

R code	data<- read.csv("D:/data/test2.csv", header=T)
	test.quan(dataset=data, group=2, quantile=281.7)

Output	P	SE_P	COVV_P			Chi	P_Value
	0.740	0.0246	0.0006	0.0002	0.0001	1.220	0.543
	0.747	0.0134	0.0002	0.0002	0.0001		
	0.760	0.0116	0.0001	0.0001	0.0001		