# DOCTORAL THESIS

## Reliable techniques for survey with sensitive question

Wu, Qin

*Date of Award:*
2013

# Reliable Techniques For Survey With Sensitive Question

**WU Qin**

**A thesis submitted in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Principal Supervisor: Dr. Tang Manlai**

**Hong Kong Baptist University**

**August 2013**

# Abstract

Refusal and providing untruthful answers from respondents are very common in the survey of collecting highly personal or sensitive issues. The randomized response model is the first interview technique designed to eliminate the non-response and response bias when sensitive questions are being asked. The idea of this technique is to hide the real status of a respondent's answer by a deliberate contamination of the data via a randomizing device. However the application of this technique is restricted by its low efficiency and high cost induced by the introduction of the randomizing device.

Non-randomized response (NRR) models have recently been developed for analyzing sensitive questions. Unlike traditional randomized response models, non-randomized response models do not require randomizing device which limits the survey format and the reproductivity of results. On the other hand, non-randomized response models introduce a non-sensitive question with binary outcome in the questionnaire in order to protect privacy and encourage cooperation from respondents. Unfortunately, the proportion of subjects who possess the non-sensitive characteristic is assumed to be known, and the non-sensitive and sensitive questions are assumed to be independent. However the two assumptions might be invalid in practical application and it can be shown that an incorrect prior knowledge about the non-sensitive question could lead to severely biased estimate of the sensitive proportion. Furthermore it is assumed that the respondents will provide their true answers applying with the survey design, but it is reported in various experimental studies that some respondents still choose to provide untruthful answers to demonstrate their positive image when they belong to the sensitive group. In this thesis several non-randomized responde techniques will be proposed to solve the known-$p$, independence assumption and noncompliance problem.

For the item count technique, it is noteworthy that all respondents only need to report the total number of statements that apply without telling which specific statements apply to them. This technique prohibits interviewers from getting the

information of the sensitive question at the level of individual subject. Despite the various attractive advantages, existing item count techniques possess three severely fatal drawbacks. Firstly, if an item count technique design consists of $m$ nonsensitive questions and one sensitive question, respondents who give an answer of $m + 1$ (i.e., answers to all questions are "yes") will automatically expose their sensitive characteristic. In this case, these respondents may refuse to answer or provide untruthful answers. This definitely weakens the applicability and credibility of the technique. Secondly, the proportion estimate and confidence interval provided in the existing papers may not be reliable as they sometimes lie outside the interval $[0, 1]$. Thirdly, the Wald-type confidence interval obtained from the technique may not be reliable for small sample designs and more reliable confidence intervals are desired. A new technique called modified item count technique is introduced to solve the above problems in which the $m$ binary non-sensitive questions are replaced by one single question which follows Poisson distribution. This technique not only protects the privacy of respondents who possess the sensitive characteristic but also provides reliable proportion estimate, shorter confidence interval, and useful sample size formulas.

**Keywords:** randomized response model, non-randomized response model, triangular model, item count technique, noncompliance.

# Table of Contents