

Survey Sampling
Statistics 4234/5234 — Fall 2017
First in-class exam

October 12, 2017

NAME:

UNI:

Instructions: Write your name and UNI in the spaces provided above, and on your bluebook. Do not turn over this page until instructed to do so.

You have 75 minutes to complete this examination. Read each part of each question carefully. There are a total of 50 points on this exam — you are responsible for checking that your paper is complete. You are permitted one $8\frac{1}{2}$ by 11 sheet (both sides) of original handwritten notes, and a hand-held calculator. **No other outside material or assistance is permitted.**

Please sign below to indicate your agreement with the Columbia College Honor Code, whether or not you are a student of Columbia College.

I affirm that I will not plagiarize, use unauthorized materials, or give or receive illegitimate help on assignments, papers, or examinations. I will also uphold equity and honesty in the evaluation of my work and the work of others. I do so to sustain a community built around this Code of Honor.

SIGNATURE: _____

Section 1: True or false? Circle the appropriate choice (1 point each).

1. The terms undercoverage and overcoverage describe disagreement between the target population and the set of units included in the sampling frame.

TRUE

FALSE

2. Convenience sampling is an unbiased sampling method, as the units that are easiest to select or most likely to respond are systematically no different than the harder-to-select or nonresponding units.

TRUE

FALSE

3. As long as a questionnaire is otherwise well designed, question wording and question ordering will have very little effect on the responses obtained.

TRUE

FALSE

4. In designing survey questionnaires, it is advisable to write specific rather than general questions, and offer a selection of choices to answer rather than simply “agree or disagree.”

TRUE

FALSE

5. In designing survey questionnaires, it is generally recognized that the more questions included, the better; if we’re going to conduct the survey, we should learn everything we can about the respondents while we have their attention.

TRUE

FALSE

6. The usual “margin of error” reported in published surveys only accounts for sampling error; the assessment of other sources of error (such as selection bias or measurement error) generally requires additional information about how the survey was conducted.

TRUE

FALSE

7. It is conceivable that estimates based on sample surveys might be *more* accurate than those based on a census, because investigators can be more careful when collecting data.

TRUE

FALSE

8. Simple random sampling is probabilistically equivalent to drawing n balls at random from an urn containing N balls; the result is that each of the $\binom{N}{n}$ possible samples has the same chance of being the sample selected.

TRUE

FALSE

9. Under simple random sampling, every unit in the population has the same probability of being selected; for an SRS of size n from a population of size N this probability is given by $\pi_i = n/N$ for $i = 1, \dots, N$.

TRUE

FALSE

10. In estimating a numerical property of a finite population based on probability sampling, an unbiased estimator is always preferable to a biased one.

TRUE

FALSE

11. Under simple random sampling, holding the sample size n fixed, the larger the population size N the more precise the estimator (in terms of lower standard error), owing to the finite population correction.

TRUE

FALSE

12. A simple random sample *with* replacement will generally yield more precise estimation of quantities of interest than ordinary SRS (without replacement).

TRUE

FALSE

13. Under simple random sampling, the true coverage rate of a nominal 95% confidence interval is equal to the percentage of the $\binom{N}{n}$ possible samples for which the CI formula, applied to that sample, would contain the true parameter value.

TRUE

FALSE

14. In estimation based on simple random sampling, the larger the sample size n , the more precise the estimator (in terms of a lower standard error) and the narrower will be the confidence interval.

TRUE

FALSE

15. As long as the sample size is sufficiently large, measurement bias, selection bias and non-sampling error are essentially non-issues, and can be ignored.

TRUE

FALSE

End of Section 1.

Section 2: Answer all questions in your bluebook.

1. (5 points) The article “What People Buy from Fast-Food Restaurants: Caloric Content and Menu Item Selection” (Obesity [2009]; 1369-1374) reported that the average number of calories consumed at lunch in New York City fast food restaurants was 827.

The researchers selected 267 fast food locations at random. The paper states that at each of these locations “adult customers were approached as they entered the restaurant and asked to provide their receipt when exiting and to complete a brief survey.”

- (a) How might nonresponse bias manifest itself in this study?
- (b) Might measurement bias be an issue? If not, why not; and if so, how?

Keep it short! One sentence on each is plenty.

2. (15 points) Consider the following population of $N = 8$ units:

i	1	2	3	4	5	6	7	8
y_i	9	12	14	13	9	15	12	12

You are given that the population mean and variance are $\bar{y}_U = 12$ and $S^2 = 4.57$, respectively.

- (a) Consider a sampling scheme in which the sample consists of either the first 4, 5 or 6 units, with equal probability. That is,

sample	probability
$\{1, 2, 3, 4\}$	$1/3$
$\{1, 2, 3, 4, 5\}$	$1/3$
$\{1, 2, 3, 4, 5, 6\}$	$1/3$

Find the sampling distribution of the sample mean \bar{y} .

- (b) For the sampling scheme in part (a), find

(i) $E[\bar{y}]$; (ii) $V[\bar{y}]$; (iii) $\text{Bias}(\bar{y})$; (iv) $\text{MSE}(\bar{y})$.

- (c) For a simple random sample of size $n = 4$, find

(i) $E[\bar{y}]$; (ii) $V[\bar{y}]$; (iii) $\text{Bias}(\bar{y})$; (iv) $\text{MSE}(\bar{y})$.

3. (15 points) In a simple random sample of 935 assistant nurses from a Norwegian county with 2700 assistant nurses, a total of 745 assistant nurses responded; 149 of the 745 respondents reported that bullying occurred in their department.

- (a) Estimate the proportion of assistant nurses in the county who would report bullying in their department.

- (b) Compute a standard error for your point estimate in part (a).

- (c) Give a 95% confidence interval for the total *number* of assistant nurses in the county who would report bullying in their department.

- (d) What assumptions must you make about the nonrespondents for the above analysis to be valid?