

36-303 Sampling, Surveys & Society

Homework 01 Solutions

February 10, 2008

1 Question 1

See Table 1.

A potential problem with this survey are the cross country comparisons. Although there exists a common protocol document to ensure a consistent application, the actual implementation of the survey within each country depends on the local team. This may raise suspicions about the quality and comparability of the data.

The technical description given in annex 1 of the main document is a little “light”. The document points to the main survey protocol document

http://www.hbsc.org/downloads/Protocol_Section1.pdf

for further details, but this document is not accessible.

2 Question 2

See Table 2. Credit to Grace DeForest.

The information we actually have about the design and application of this survey is minimal. We do not know basic things such as how the respondents were selected or how the instrument was administered or which the sampling frame was. All the information available comes from press releases and there are not any accessible technical documents. Without this information the results of this survey are hard to believe.

Table 1: Survey: Health Behavior in School-Age Children Study

Sponsor	Cross national effort with the collaboration of the World Health Organization
Collector	Collected by each participant country by an HBSC team and all countries data gathered and stored at the Norwegian Social Science Data Services.
Purpose	Gain new insight into, and increase understanding of young people's health and well-being, health behaviours and their social context. The results must be allow inter-country comparisons.
Year started	1982
Target Population	Young people attending school, aged 11, 13 and 15 years old.
Sampling Frame	Varies per country. Usually a list of classes. A list of schools if a list of classes is not available.
Sample Design	Clustered probability design. For each country classes (or schools) are the PSU and the students are randomly selected within them.
Sample Size	Varies per country. A total of 162,306 samples were collected counting all countries.
Use of Interviewer	None. Self administered. Supervised by teachers. In some countries, a team of trained interviewers was employed.
Mode of Administration	Self administered.
Computer Assistance	None.
Reporting Unit	Student
Time dimension	Administered over a 7-8 month period from October to May.
Frequency	Every four years.
Interviews per Round of Survey	One
Levels of observation	Countries, classes (or schools where not available) and students.
Web Link	http://www.hbsc.org/

Table 2: Survey: Hiremenow.com

Sponsor	Hireme.com
Collector	Hireme.com
Purpose	- To determine if employers are willing to watch video resumes. - To determine the appropriate length of a video resume.
Year started	2007
Target Population	Employers seeking temporary and contract-work employees.
Sampling Frame	300 human resource and business managers
Sample Design	Not clear
Sample Size	300 subjects
Use of Interviewer	Not clear.
Mode of Administration	Not clear.
Computer Assistance	Not clear
Reporting Unit	Selected employers
Time dimension	administered over a one-month period
Frequency	Single application
Interviews per Round of Survey	not clear
Levels of observation	Human resources and business managers. Not clear about other levels.
Web Link	www.hiremenow.com

3 Question 3

3.1 (a)

$$\begin{aligned}E[X_i] &= E[X_1] \quad (\text{because the } X_i\text{'s are iid}) \\&= 1 \cdot p + 0 \cdot (1 - p) \\&= p\end{aligned}$$

$$\begin{aligned}V[X_i] &= E[X_i^2] - E[X_i]^2 \\&= E[X_1^2] - E[X_1]^2 \\&= 1^2 \cdot p + 0^2 \cdot (1 - p) - p^2 \\&= p - p^2 = p(1 - p)\end{aligned}$$

3.2 (b)

$$\begin{aligned}E \left[\sum_{i=1}^n X_i \right] &= \sum_{i=1}^n E[X_i] \\&= \sum_{i=1}^n E[X_1] \\&= np\end{aligned}$$

$$\begin{aligned}V \left[\sum_{i=1}^n X_i \right] &= \sum_{i=1}^n V[X_i] \\&= \sum_{i=1}^n V[X_1] \\&= np(1 - p)\end{aligned}$$

Note that the identity $V \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n V[X_i]$ is true because we know that the random variables $X_1, X_2, \dots, X_n$ are independent. This is **not** true in general.

3.3 (c)

Using the results from the previous sections,

$$\begin{aligned}E[\hat{p}] &= E\left[\frac{Y}{n}\right] \\&= \frac{1}{n}E[Y] \\&= \frac{1}{n} \cdot np \\&= p\end{aligned}$$

The above property is called "unbiasedness". In this case we say that $\hat{p}$ is an *unbiased* estimator of p .

Again using the results from the previous sections,

$$\begin{aligned}V[\hat{p}] &= V\left[\frac{Y}{n}\right] \\&= \frac{1}{n^2}V[Y] \\&= \frac{1}{n^2} \cdot np(1-p) \\&= \frac{p(1-p)}{n}\end{aligned}$$