36-303 Sampling, Surveys & Society
Homework 01 Solutions

February 10, 2010

1 Question 1

See Table 1.

(b) I couldn't find any information about computer assistance, although I think they probably used computer assistance in randomizing geographic points and telephone numbers. Also, it is simply stated that one student from each household was selected by a systematic procedure to provide an approximate balance of respondents by sex, readers would like to know more details about the systematic procedure to decide on the validity of the survey. It is generally common to get nonresponse in telephone survey, I couldn't find information about how was this problem dealt with in the survey so there might be some bias in the findings. I am not sure if the stratification method used in this survey was enough to draw conclusions about complex issues such as diversity of America's youth.

(c)

$$\frac{2 * \sqrt{0.5 * 0.5}}{\sqrt{1005}} = 3.15\%$$

2 Question 2

See Table 2.

(b) The information we actually have about the design and application of this survey is minimal. We do not know basic information such as the sampling frame and sampling design.

(c)

$$\frac{2 * \sqrt{0.5 * 0.5}}{\sqrt{302}} = 5.75\%$$

Table 1: Survey: The State of Our Nation's Youth

Sponsor	The Horatio Alger Association of Distinguished American
Collector	collected by Peter D. Hart Research Associates
Purpose	to give voice to the issues, opinions and thoughts of our nation's young people.
Year started	2005
Target Population	students age 13 to 19 who identified themselves as ninth-through-twelfth graders
Sampling Frame	compiled list provided by american student list, the well-respected national list management firm, which specializes in maintaining lists of k-12 students
Sample Design	sample was created by selecting 505 geographic points randomly and proportionate to the population of each region and, within each region, by size of place. All telephone numbers have an equal chance to be included. One student from each selected household was included, selected by a systematic procedure to provide a balance of respondents by sex
Use of Interviewer	interviewer administered
Mode of Administration	telephone interview
Computer Assistance	not clear
Reporting Unit	student
Time dimension	conducted from May 9 to 12, 2005
Frequency	every year
Interviews per Round of Survey	one
Levels of observation	student
Web Link	`http://www.horatioalger.com/pdfs/state05.pdf/`

Table 2: Survey: Raising the Bar: Employers' views on college learning in the wake of the economic downturn

Sponsor	The Association of American Colleges and Universities
Collector	hart Research Associates
Purpose	(1)find out if the majority of employers are expect to increase their hiring in the next year. If so, what kind of education background are most favorable to them for hiring (2) find out what kind of skills for new graduates do employers favor for hiring (3)find out if the employers are satisfied with college today in educating their employees to face challenges. If not, what needs to be improved in their opinion
Year started	2009
Target Population	employers whose organizations have at least 25 employees and report that 25% or more of their new hires hold either an associates degree from a two-year college or a bachelor's degree from a four-year college
Sampling Frame	not clear
Sample Design	not clear
Sample Size	302 employers
Use of Interviewer	yes, by associates at Hart research
Mode of Administration	not clear
Computer Assistance	not clear
Reporting Unit	selected executives at private sector and non-profit organizations, including owners, CEOs, presidents, C-suite level executives, and vice presidents
Time dimension	conducted from October 27 to November 17,2009
Frequency	single application
Interviews per Round of Survey	one
Levels of observation	employer, company
Web Link	`http://www.aacu.org/leap/documents/2009_EmployerSurvey.pdf`

3 Question 3

3.1 (a)

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

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

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 $X_1, X_2, ..., X_n$ 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,

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

4 Question 4

4.1 (a)

$$\begin{aligned} P[X=1] = P[X=1, Y=4] + P[X=1, Y=3] &= \frac{1}{2} \\ P[X=2] = 1 - P[X=1] &= \frac{1}{2} \\ P[Y=3] = P[X=1, Y=3] + P[X=2, Y=3] &= \frac{1}{2} \\ P[Y=4] = 1 - P[Y=3] &= \frac{1}{2} \end{aligned}$$

Now, to prove that X and Y are not independent, we just need to show a counter example:

$$(P[X=1, Y=4] = \frac{1}{8}) \neq (P[X=1] * P[Y=4] = \frac{1}{4})$$

4.2 (b)

assuming X and Y are independent, then

$$P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{P[X = x] * P[Y = y]}{P[Y = y]} = P[X = x]$$