Stat 572 Sampling Theory & Practice
Homework 5

Erik Barry Erhardt

April 18, 2006

Assignment 5.1 *Board defects.*

(a) *Sample four.*
Table 1 on page 1 gives the probabilities of selecting each of the 10 boards.

Board i	ϕ_i
1	10/150=0.067
2	12/150=0.080
3	22/150=0.147
4	8/150=0.053
5	16/150=0.107
6	24/150=0.160
7	9/150=0.060
8	10/150=0.067
9	8/150=0.053
10	31/150=0.207

Table 1: 5.1 Probability of selecting each of the 10 boards.

There are many ways of drawing such a sample. One way is draw a sample of size 4 from a multinomial distribution, with bin i having probability equal to ϕ_i. If any bin is over sampled (more drawn than there are individuals, more than one in this case), then redraw the sample. This method does not require sequential draws with probability updates.

(b) *Average number of defects per board.*

Estimates are first calculated in terms of totals, then corrected for averages per board.

$N = 10$ boards.

$$\begin{aligned}
\hat{\bar{y}}_\phi &= \frac{1}{N}\hat{t}_\phi \\
&= \frac{1}{10}\frac{1}{n}\sum_{j=1}^{n}(\hat{t}_\phi)_j \\
&= \frac{1}{10}\frac{1}{4}\left(1/(12/150) + 3/(22/150) + 2/(16/150) + 1/(9/150)\right) \\
&= \frac{1}{10}\frac{1}{4}\left(12.5000 + 20.4545 + 18.7500 + 16.6667\right) \\
&= \frac{1}{10}17.0928 \\
&= 1.70928 \\
\mathrm{Var}(\hat{\bar{y}}_\phi) &= \frac{1}{N^2}\mathrm{Var}(\hat{t}_\phi) \qquad [\text{Var has } n \text{ in denominator instead of } n-1] \\
&= \frac{1}{100}8.8308 \\
&= 0.088308
\end{aligned}$$

$\hat{\bar{y}}_\phi = 1.70928$ and $\mathrm{SE}(\hat{\bar{y}}_\phi) = \sqrt{0.088308} = 0.2972$.

Assignment 5.2 *Randomized response.*

$\theta = 0.80$ is probability of card A "Have shots" and $1 - \theta = 0.20$ is probability of card B "Not shots". $n = 200$ SRS, with $n_1 = 145$" "yeses".

$$\begin{aligned}
\hat{p} &= \frac{n_1/n - (1-\theta)}{2\theta - 1} \\
&= \frac{145/200 - 0.2}{1.6 - 1} \\
&= \frac{0.525}{0.6} \\
&= 0.875 \\
\text{SE}(\hat{p}) &= \sqrt{\left(\frac{1}{2\theta - 1}\right)^2 \left(\frac{1}{n}\right) \left(\frac{n_1}{n}\right) \left(\frac{n - n_1}{n}\right)} \\
&= \sqrt{\left(\frac{5}{3}\right)^2 \left(\frac{1}{200}\right) \left(\frac{145}{200}\right) \left(\frac{55}{200}\right)} \\
&= \sqrt{0.002769} \\
&= 0.0526
\end{aligned}$$

95% CI: 0.875 ± 0.103 gives $(0.772, 0.978)$.

Assignment 5.3 *Certified Statisticians.*

(a) *Response rates.*

Table 2 on page 4 gives the response rates from the ASA population of size $N = 18609$.

Group	Pop prop	N_i	n_i	Response Rate $\frac{n_i}{N_i}$
P	0.55	10235	3036	0.2966
M	0.38	7071	1640	0.2319
Totals	0.93	17306	4676	
I	0.29	5397	1809	0.3352
A	0.34	6327	2221	0.3510
G	0.11	2047	880	0.4299
Totals	0.74	13771	4910	

Table 2: 5.3 Response Rates.

To test whether the nonresponses are missing completely at random (MCAR) among college level, I run a chi-square test to see whether the response rates by group are the same. The table in Figure 3 on page 5 gives a $X^2 = 88.765$ with $df = 1$ for a p-value $= 0.000$. Thus, there is strong evidence that the nonresponses are not missing COMPLETELY at random. They may be missing at random (MAR), but to tell whether the nonreponses are ignorable (MAR) or nonignorable, we would need to resample the nonresponses and see whether the responses are different between the respondants and the "responding nonrespondants".

To test whether the nonresponses are missing completely at random (MCAR) among work environment, I run a chi-square test to see whether the response rates by group are the same. The table in Figure 4 on page 5 gives a $X^2 = 59.577$ with $df = 2$ for a p-value $= 0.000$. Thus, there is strong evidence that the nonresponses are not missing COMPLETELY at random.

As it is often the case that those with the strongest opinions self-select themselves to respond to surveys, it is unlikely that these responses are MAR.

```
Chi-Square Test: R1, N1

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

           R1       N1  Total
    1    3036     7199  10235
      2765.45  7469.55
       26.469    9.799

    2    1640     5431   7071
      1910.55  5160.45
       38.312   14.184

Total    4676    12630  17306

Chi-Sq = 88.765, DF = 1, P-Value = 0.000
```

Table 3: 5.3a1 Chi-square test of equal response rates concludes they are different.

```
Chi-Square Test: R2, N2

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

           R2       N2  Total
    1    1809     3588   5397
      1924.28  3472.72
        6.906    3.827

    2    2221     4106   6327
      2255.87  4071.13
        0.539    0.299

    3     880     1167   2047
       729.85  1317.15
       30.890   17.116

Total    4910     8861  13771

Chi-Sq = 59.577, DF = 2, P-Value = 0.000
```

Table 4: 5.3a2 Chi-square test of equal response rates concludes they are different.

(b) *Favoring certification.*

Using a poststratification on the work environment, I calculate the estimate of the proportion of ASA membership favoring certification (response = 1).

$$
\begin{aligned}
\hat{p}_{\text{post}} &= \sum_{\ell=1}^{L} \frac{N_\ell}{N}\hat{p}_{\ell R} \\
&= \frac{5397}{13771}(0.2963) + \frac{6327}{13771}(0.25394) + \frac{2047}{13771}(0.2250) \\
&= 0.26624 \\
S_\ell^2 &= \frac{\hat{p}_{\text{post}}(1-\hat{p}_{\text{post}})}{n_{\ell R}-1} \\
\text{Var}(\hat{p}_{\text{post}}) &= \sum_{\ell=1}^{L} \left(\frac{N_\ell}{N}\right)\left(\frac{N_\ell - n_{\ell R}}{N_\ell}\right) S_\ell^2 \\
&= 0.000025965
\end{aligned}
$$

The estimate is $\hat{p}_{\text{post}} = 0.26624$ with $\text{SE}(\hat{p}_{\text{post}}) = 0.0050956$.

(c) *Majority opposes?*

It is unlikely that any of the ASA members would take part (b) as evidence that a majority opposes certification. There are more Nos than Yeses, however. Calculating the estimate of the proportion of ASA membership opposing certification (response = 5). The estimate is $\hat{p}_{\text{post}} = 0.3913$ with $\text{SE}(\hat{p}_{\text{post}}) = 0.005605$.

Yet, considering both yes and possibly responses (response = 1 or 2). The estimate is $\hat{p}_{\text{post}} = 0.4860$ with $\text{SE}(\hat{p}_{\text{post}}) = 0.0057483$.

Further, considering the Table 5 on page 6, there are a few more who favor certification than do not favor, though those who do not favor certification tend to feel stronger about it (more Nos to Unlikelies, than Yeses to Possiblies).

Value	Count	Percent
No Response 0	12	0.24%
Yes 1	1321	26.41%
Possibly 2	1114	22.28%
No Opinion 3	269	5.38%
Unlikely 4	337	6.74%
No 5	1948	38.95%

Table 5: 5.3c

Appendix

code used for the above analysis

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 5.1
x=[10 12 22 8 16 24 9 10 8 31]
p=x/sum(x)
[[1:10]' x' sum(x)*ones(length(x),1) p']
c=[2 3 5 7];
d=[1 3 2 1];
p=[12/150 22/150 16/150  9/150];
that=mean(d./p)
ybar=that/10
var_that = var(d./p,1)
var_ybar = var_that/(10^2)
se_ybar = sqrt(var_ybar)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 5.2
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 5.3
[xn,xc]=xlsread('certify.xls');
college=xc([2:end],7);
workenv=xc([2:end],9);
tabulate(college)
tabulate(workenv)
% Minitab
R1 N1    R2 N2
3036 7199 1809 3588
1640 5431 2221 4106
*     *        880   1167

(c)
% Yeses
p=[.25394 .2250 .2963]
N=[6327 2047 5397]
phat = sum(N.*p/sum(N))
n=[2221 880 1809]
S2=p.*(1-p)./(n-1)
var_ypost = sum(((N./sum(N)).^2).*(1-n./N).*(S2))
%Nos
p=[.3926 .4273 .3720]
phat = sum(N.*p/sum(N))
S2=p.*(1-p)./(n-1)
var_ypost = sum(((N./sum(N)).^2).*(1-n./N).*(S2))
%Yes+Poss
p=[.4818 .4386 .5141]
phat = sum(N.*p/sum(N))
S2=p.*(1-p)./(n-1)
var_ypost = sum(((N./sum(N)).^2).*(1-n./N).*(S2))

tabulate(xn(:,1))
% EOF
```