Stat 572 Sampling Theory & Practice
Homework 2

Erik Erhardt

February 9, 2006

All relevant code is included in the appendix.

Assignment 2.1

	Population Mean	Population Proportion
$CI_{\text{with finite pop. correction}}$:	$\bar{y}_s \pm z_{\frac{\alpha}{2}} \sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}$	$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\left(\frac{N-n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1}}$
$CI_{\text{without finite pop. correction}}$:	$\bar{y}_s \pm z_{\frac{\alpha}{2}} \sqrt{\frac{s^2}{n}}$	$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$

When the finite population correction $fpc = 0.99746$ is applied with a population size of $N = 19664$ and sample size $n = 50$, there is almost no difference in the shrinkage of the confidence intervals.

(a) *Q4, average price willing to pay for cable TV*

$\bar{y}$	95% CI	finite pop. correction
9.8	(8.0103, 11.59)	with
9.8	(8.008,11.592)	without

(b) *Q2 proportion of households having at least one child aged 11 or younger*

$\hat{p}$	95% CI	finite pop. correction
0.36	(0.34102,0.37898)	with
0.36	(0.34099,0.37901)	without

Assignment 2.2 *Sample size calculation*

Given a bound of $B = \$0.50$, a 95% confidence level for $z_{\frac{\alpha}{2}} = 1.960$, population size $N = 19664$, and by estimating the population standard deviation S with the sample standard deviation $s = 6.465$, we can solve for n using $n = \frac{z^2_{\frac{\alpha}{2}} s^2}{B^2 + z^2_{\frac{\alpha}{2}} s^2/N} = 621.92$. Rounding up gives a sample size of $n = 622$.

Assignment 2.3 The margin of error below is the desired $B = \$0.50$.

(a)

$\bar{y}$	95% CI	finite pop. correction
9.6463	(9.1423,10.1503)	with
9.6463	(9.1342,10.1584)	without

Margin of error = 0.5040

(b)

$\hat{p}$	95% CI	finite pop. correction
0.3424	(0.3410,0.3439)	with
0.3424	(0.3409,0.3439)	without

Margin of error = 0.0015

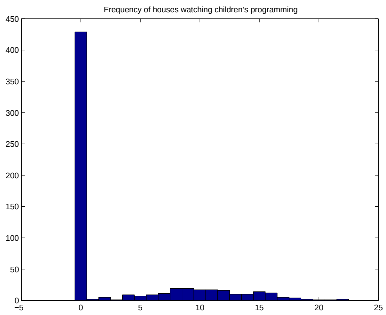
Assignment 2.4

$\bar{y}$	95% CI	finite pop. correction
70367	(69054,71680)	with

The CI includes \$71,117.

Assignment 2.5 Because there are many households that do not watch any children's programming, this distribution is not close to symmetric because of a point mass at 0 hours. For the houses that do watch children's programming, the distribution is roughly symmetric about 10.5.

In order to use the normal approximation for a CI for the mean, the underlying distribution should be roughly symmetric, or the sample size should be large enough to overcome non-normality from the underlying distribution so that the distribution of the mean is roughly normal. A sample size of $n = 622$ is very large, so there should be little problem using the normal approximation for the CI for the population mean, even with the extreme skewness.


Frequency of houses watching children's programming
450
400
350
300
250
200
150
100
50
0
−5
0
5
10
15
20
25

Appendix

code used for the above analysis

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% 2.1
./addgen
hw2_add1
27511
51-75
50
no
./survey
hw2_add1
hw2_sam1
0 0 0

% edit hw2_sam1 removing first and last line, and saving hw2_sam1.txt
% matlab
x=load('hw2_sam1.txt');

y=[x(:,7), ones(length(x),1).*(x(:,5)>0)];   % convert to binary variable

N=19664;n=length(x);
alpha=0.05;
mu = [mean(y(:,1)),mean(y(:,2))]
s  = [std(y(:,1)), sqrt(mu(2)*(1-mu(2))/(n-1))]
fpc = (N-n)/N                                % finite population correction
me = norminv(1-alpha/2,0,1)*sqrt(fpc*s.^2/n)
CI = [mu - me; mu + me]
[mu' CI']
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% without finite population correction
me = norminv(1-alpha/2,0,1)*sqrt(s.^2/n)
CI = [mu - me; mu + me]
[mu' CI']
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% 2.2
B=0.50;
n=(norminv(1-alpha/2,0,1)^2*s(1)^2)/(B^2+norminv(1-alpha/2,0,1)^2*s(1)^2/N)
n_samsize=ceil(n)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% 2.3
./addgen
hw2_add2
71008
51-75
622
no
./survey
hw2_add2
hw2_sam2
0 0 0

% edit hw2_sam2 removing first and last line, and saving hw2_sam2.txt
% matlab
x=load('hw2_sam2.txt');

y=[x(:,7), ones(length(x),1).*(x(:,5)>0)];   % convert to binary variable

N=19664;n=622;
alpha=0.05;
mu = [mean(y(:,1)),mean(y(:,2))]
s  = [std(y(:,1)), sqrt(mu(2)*(1-mu(2))/(n-1))]
fpc = (N-n)/N                                % finite population correction
me = norminv(1-alpha/2,0,1)*sqrt(fpc*s.^2/n)
CI = [mu - me; mu + me]
[mu' CI']
me
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% without finite population correction
me = norminv(1-alpha/2,0,1)*sqrt(s.^2/n)
CI = [mu - me; mu + me]
[mu' CI']
me
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% 2.4

N=19664;n=622;
alpha=0.05;
mu = mean(x(:,[3]))
s  = std(x(:,[3]))
fpc = (N-n)/N                                % finite population correction
```

```
me = norminv(1-alpha/2,0,1)*sqrt(fpc*s.^2/n)
CI = [mu - me; mu + me]
[mu' CI']

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% 2.5
[a1,a2]=hist(x(:,11),[0:22])
[a2;a1]'

hist(x(:,11),[0:22])
title('Frequency of houses watching children''s programming')
```