Referred Solution to Assignment 02

Part I : Problems from Chapter 2

2.12. (a) $n_0 = \left(\dfrac{z_{\alpha/2}\sqrt{p(1-p)}}{e}\right)^2 = \dfrac{1.96^2 \cdot 0.5 \cdot (1-0.5)}{0.1^2} = 96.04 \Rightarrow n_0 = 96$ or 97.

If we take $n_0 = 97$, after the fpc adjustment, $n = \dfrac{n_0}{1+\dfrac{n_0}{N}} = \dfrac{97}{1+\dfrac{97}{580}} = 83.1$.

So, we need to take sample size 84.

(b) Since a SRS with replacement is adopted, we do not need to consider fpc. Let $\hat{p} = \dfrac{27}{120}$, 95% confidence interval for p:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \Rightarrow (0.150, 0.300).$$

2.18. If we refer to 2.16, we have the population size 14938. However, $1 - \dfrac{n}{N} \approx 1$. Therefore, in this case, we may form the 95% CI for the population proportion without fpc.

Let $\hat{p} = 0.7083$, 95% confidence interval for p:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \Rightarrow (0.627, 0.790).$$

2.19. $n_0 = \left(\dfrac{z_{\alpha/2}\sqrt{p(1-p)}}{e}\right)^2 = \dfrac{1.96^2 \cdot 0.5 \cdot (1-0.5)}{0.04^2} = 600.25 \Rightarrow n_0 = 600$ or 601.

If we take $n_0 = 601$, after the fpc adjustment, the sample size for each city should be as follows:

City	Population Size (N)	Sample Size (n)
Buckeye	4857	535
Gilbert	59338	595
Gila Bend	1724	446
Phoenix	1149417	601
Tempe	153821	599

The fpc would make differences for cities, Buckeye and Gila Bend, because both of them have small population sizes.

2.20. In this case, under the fpc, we have $SE(\hat{t}_s) = \sqrt{\left(1-\dfrac{n}{N}\right)} \cdot \dfrac{8s}{\sqrt{n}}$ and use the attached R-code offered by Lynetta Campbell. We obtain the confidence level $\dfrac{60}{70} = 85.7\%$ which does not equal 95%.

```
# Enter the data
x = c(1:8)
y = c(1,2,4,4,7,7,7,8)
n = length(y)
n
t_true = sum(y)
## Write a function to build the required table:
fun = function() {
cat("Sample","     ","y_i","    ","t_hat","  ","SE","      ",
"Conf. Int.","         ","u(S)","\n")
cntr = 0
for(i in 1:5){
for(j in (i+1):6) {
for(k in (j+1):7) {
for(m in (k+1):8) {
zip = c(i,j,k,m)
y_s = y[zip]
t_s = 8*mean(y_s)
s_s = sd(y_s)
SE = round(8*((s_s^2/4)*(1-4/8))^0.5,2)
t_lo = t_s-1.96*SE
t_hi = t_s+1.96*SE
if(t_true>t_lo & t_true<t_hi) u_S=1 else u_S = 0
cat(zip,"   ",y_s,"   ",t_s,"   ",SE,"  ",
"[",t_lo,",",t_hi,"]","   ",u_S, "\n")
cntr = cntr + u_S
}}}}
cat("\n")
cat("Number of CIs that included 40:","\n")
return(cntr)
}
fun()
```

Part II : Extra Problems

1. Since both plans are random samples, theoretically, the estimates of average income are unbiased. Moreover, the two countries have roughly similar wealth distributions and we assume both population standard deviation (S) are known. Then, we have the standard error of the estimate of average income for each country as follows:

$$SE_{\text{A}} = \sqrt{\left(1 - \frac{100}{1000}\right)} \frac{S}{\sqrt{100}} \approx 0.095S,$$

$$SE_{\text{B}} = \sqrt{\left(1 - \frac{1000}{100000000}\right)} \frac{S}{\sqrt{1000}} \approx 0.032S.$$

Therefore, in country B, the statistician will get a better estimate of average income since it has smaller variance for mean estimate.

2. (a) Assume the election results is the true proportion (p) and the prediction is $\hat{p}$. First, we evaluate the likelihood of the sampling error over $\pm 4.4\%$ in 1952 and the likelihood of the sampling error over $\pm 2.7\%$ in 1964.

$$\begin{aligned}\text{In 1952,}\quad & P(\hat{p}-p<-0.044)+P(\hat{p}-p>0.044)\\ &=2\times P(|\hat{p}-p|>0.044)\\ &=2\times P\left(|Z|>\frac{0.044}{\sqrt{\frac{(0.51)(0.49)}{5384}}}\right)\\ &\approx 1.0585\times 10^{-10}.\end{aligned}$$

$$\begin{aligned}\text{In 1964,}\quad & P(\hat{p}-p<-0.027)+P(\hat{p}-p>0.027)\\ &=2\times P(|\hat{p}-p|>0.027)\\ &=2\times P\left(|Z|>\frac{0.027}{\sqrt{\frac{(0.64)(0.36)}{6624}}}\right)\\ &\approx 4.6928\times 10^{-6}.\end{aligned}$$

Second, to evaluate the precision of prediction of the election results, we use the marginal of error to form the confidence intervals for the election results.

$$\text{In 1952, 95\% CI for } p: 0.51\pm 1.96\sqrt{\frac{(0.51)(0.49)}{5384}} \Rightarrow (0.497, 0.523).$$

$$\text{In 1964, 95\% CI for } p: 0.613\pm 1.96\sqrt{\frac{(0.64)(0.36)}{6624}} \Rightarrow (0.628, 0.652).$$

From the above discussions, we conclude that in 1952, the likelihood of the sampling error over $\pm$ 4.4% is very tiny; however, under the 95% confidence level, the confidence interval does not capture the true election result. Therefore, we may concern the precision of prediction of the election result. Similarly, we have the same conclusion in 1964.

(b)

Year	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992
$\hat{p}-p$	-0.044	0.017	0.009	0.027	-0.005	0.002	-0.016	-0.037	-0.002	0.021	0.058

Theoretically, the bias is $E(\hat{p}-p)$. So, the estimated bias from the poll over 1952-1992 can be $\frac{-0.044+0.017+\cdots+0.058}{11}\approx +0.273\%$.