3430 Assignment 3

Ravish Kamath: 213893664

14 March, 2022

Question 1

The personnel manager of a corporation wants to estimate, for one year, the total number of days used for sick leave among all 46 plants in his firm. The 46 plants are divided into 20 "small" plants and 26 "large" plants. From past experience, the manager figures that the small plants may use from 0 to 100 days of sick leave, whereas the large plants may use from 10 to 200 days of sick leave. If he desires to estimate the total to within 100 days:
a. Find the appropriate allocation of the sample to the two strata.
b. Find the appropriate sample size.

Solution

Part A

We are given 2 strata which are 20 small plants and 26 large plants. Hence our N_h is 20 and 26. Furthermore we are given our range for each strata which are 0 to 100 for small plants and 10 to 200 for large plants. We are unfortunately not given the cost of each observation. Recall the formula for the the sample allocation fraction:

$$a_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h}$$

We have already stated what N_h but need to figure out S_h which is $\frac{range}{4}$.

```
Nh = c(20,26)
range = c(100-0, 200-10)
Sh = range/4
```

Let us now input the these values into our sample allocation fraction eqaution.

```
ah = round((Nh*Sh)/sum(Nh*Sh),4)
ah
```

```
## [1] 0.2882 0.7118
```

Part B

To find the appropriate sample size we will use this formula:

$$\frac{(\sum_{h=1}^{H} N_h S_h)^2}{N^2 D + \sum_{h=1}^{H} N_h S_h^2}$$

Furthermore, we are given a bound of 100 days.

```
B = 100
NsqrdD = B^2/4
```

```
n = ceiling((sum(Nh*Sh)^2) / (NsqrdD + sum(Nh*(Sh^2))))
n
```

```
## [1] 41
```

```
nh = ceiling(n*ah)
nh
```

```
## [1] 12 30
```

We are given a full sample size of 41, with an allocation of 12 to small plants and 30 to the large plants. However we come accross an issue since there are only 26 large plants. Hence we will use the full 26 large plants in our sample and the rest of the small plants will be used in the sample which would be 15.

Question 2

Wage earners in a large firm are stratified into management and clerical classes, the first having 300 and the second having 500 employees. To assess attitude on sick-leave policy, independent random samples of 100 workers each were selected, one sample from each of the classes. After the sample data were collected, the responses were divided according to gender. In the table of results, a = Number who like the policy; b = Number who dislike the policy; and c = Number who have no opinion on the policy.

Find an estimate and an estimated variance of that estimate for each parameter listed:
a. Proportion of managers who like the policy
b. Proportion of wage earners who like the policy
c. Total number of female wage earners who dislike the policy
d. Difference between the proportion of male managers who like the policy and the proportion of female managers who like the policy
e. Difference between the proportion of managers who like the policy and the proportion of managers who dislike the policy

Solution

Let us first recreate the table given in the question.

```
##        Mang a Mang b Mang c Cler a Cler b Cler c
## Male       60     15      5     24      4      2
## Female     10      7      3     42     30      8
```

Further more we are given our populations and sample strata sizes which gives us our information on overall population and sample size. Finally we can also figure out the sampling fraction denoted by f_h.

```
nh = c(100,100)
n = sum(nh)
Nh = c(300,500)
N = sum(Nh)
Wh = Nh/N
fh = nh/Nh
```

Part A

Our estimated proportion, $\hat{p_h}$, of managers who like the policy is **0.7**.

```
hat_pMang = (df[1,1] + df[2,1])/100
```

Our estimated variance, $\hat{V}(\hat{p})$, is **0.0014**.

```
ssh = (nh[1]/(nh[1]-1))*(hat_pMang*(1-hat_pMang))
varhat_pMang = (1-fh[1])*(ssh/(nh[1]-1))
```

Finally our bound on error, $\mathbf{B = 0.07558904}$:

```
B = 2*sqrt(varhat_pMang)
hat_pMang + B
hat_pMang - B
```

We have a confidence interval of **(0.624411, 0.775589)**.

Part B

Our estimated proportion, $\hat{p_{st}}$ of wage earners who like the policy is **0.675**:

```
hat_pCler = (df[1,4] + df[2,4])/100
hat_ph = c(hat_pMang, hat_pCler)
hat_pSt = sum(Wh*hat_ph)
```

Our estimated variance, $\hat{V}(\hat{p_{st}})$, is 0.000907197.

```
hat_pq = (hat_ph*(1-hat_ph))
hat_var_pSt = sum((Wh^2)*(1-fh)*hat_pq/(nh-1))
```

Finally our bound on error, $\mathbf{B = 0.06023942}$:

```
B = 2*sqrt(hat_var_pSt)
hat_pSt + B
hat_pSt - B
```

We have a confidence interval of **(0.6147606, 0.735294)**.

Part C

The estimated total, $\hat{\tau_{st}}$, number of female wage earners who dislike the policy is **171**.

```
hat_pMang = df[2,2]/nh[1]
hat_pCler = df[2,5]/nh[2]
hat_ph = c(hat_pMang, hat_pCler)
hat_pSt = sum(Wh*hat_ph)
hat_tauSt = N*hat_pSt
```

Our estimated total variance, $\hat{V}(\hat{\tau_{st}})$, is **463.697**:

```
hat_pq = (hat_ph*(1-hat_ph))
hat_var_pSt = sum((Wh^2)*(1-fh)*hat_pq/(nh-1))
hat_var_tauSt = N^2 * hat_var_pSt
```

Finally our bound on error, $\mathbf{B = 43.06725}$:

```
B = 2*sqrt(hat_var_tauSt)
hat_tauSt + B
hat_tauSt - B
```

We have a confidence interval of **(127.9328, 214.0672)**.

Part D

Our estimated difference between the proportion of male managers who like the policy and the proportion of female managers who like the policy is **0.5**.

```
hat_pMangM = df[1,1]/nh[1]
hat_pMangF = df[2,1]/nh[1]
```

We have the estimated variance for Male Managers to be **0.0002272727** and for Female Managers to be **8.522727e-05**.

```
sh_sqrd = (nh[1]/(nh[1]-1))*(hat_pMangM*(1-hat_pMangM))
varhat_pMangM = Wh[1]^2*(1-fh[1])*sh_sqrd/nh[1]

sh_sqrd2 = (nh[1]/(nh[1]-1))*(hat_pMangF*(1-hat_pMangF))
varhat_pMangF = Wh[1]^2*(1-fh[1])*sh_sqrd2/nh[1]
```

We have the bound on error, **B = 0.03535534**.

```
B = 2*sqrt(varhat_pMangM + varhat_pMangF)
(hat_pMangM - hat_pMangF) + B
(hat_pMangM - hat_pMangF) - B
```

We have a confidence interval of **(0.4646447, 0.5353553)**.

Part E

Our estimated difference between the proportion of managers who like the policy and the proportion of managers who dislike the policy is **0.48**.

```
hat_pMang_a = (df[1,1] + df[2,1])/100
hat_pMang_b = (df[1,2] + df[2,2])/100
```

Here we have the estimated variance for managers who like the policy to be **0.0001988** and for managers who dislike the policy to be **0.0001625**.

```
sh_sqrd= (nh[1]/(nh[1]-1))*(hat_pMang_a*(1-hat_pMang_a))
varhat_pMang_a = Wh[1]^2*(1-fh[1])*sh_sqrd/nh[1]

sh_sqrd2 = (nh[1]/(nh[1]-1))*(hat_pMang_b*(1-hat_pMang_b))
varhat_pMang_b = Wh[1]^2*(1-fh[1])*sh_sqrd2/nh[1]
```

We have a bound of error, **B = 0.03801913**.

```
B = 2*sqrt(varhat_pMang_a + varhat_pMang_b)
(hat_pMang_a - hat_pMang_b) - B
(hat_pMang_a - hat_pMang_b) + B
```

We have a confidence interval of **(0.4419809, 0.5180191)**.

Question 3

Are anesthesiologists overworked and therefore putting patients at risk? This question was investigated as part of a survey carried out at the University of Florida. The popula- tion of those practicing anesthesiology was stratified into three groups: anesthesiologists (composing approximately 50% of the population), anesthesiology residents (composing approximately 10% of the population), and nurse anesthetists (composing approximately 40% of the population). The frequencies of those in each stratum who thought they had worked without a break beyond a safe limit sometime during the last 6 months are shown in the accompanying table:

a. Estimate the population proportion of those who think they have worked beyond a safe limit. Calculate a bound on the error of estimation.
b. Do anesthesiologists differ significantly from residents in this matter?
c. Do anesthesiologists differ significantly from nurse anesthetists in this matter?

Solution

Part A

We are given 3 strata which are anesthesiologist, anesthesiology resident and nurse anesthetist. We need to assume S_h^2 are equal to each strata, hence $a_h = W_h$. Furtheremore, we can assume the population strata size, N_h, is quite large since as the overall population is for the United States, we can ignore fpc to calculate the variance.

```
nh = c(913 + 417, 136 + 29, 860 + 240)
n = sum(nh)
ah = nh/n

yes_ph = c(.687,.824,.782)
no_ph = c(.314,.176,.218)

hat_pSt = sum(ah*yes_ph)
hat_pSt
```

```
## [1] 0.7359807
```

```
hat_pq = yes_ph*no_ph
hat_var_pSt = sum( (ah^2) * (hat_pq/(nh - 1)) )
hat_var_pSt
```

```
## [1] 7.408495e-05
```

```
B = 2*sqrt(hat_var_pSt)
B
```

```
## [1] 0.01721452
```

```
hat_pSt + B
hat_pSt - B
```

As shown above we are given the estimated population proportion to be **0.736** with a bound of error of $\mathbf{B} =$ **0.017**. This gives us a confidence of interval **(0.7187662, 0.7531953)**.

Part B

```
(yes_ph[1] - yes_ph[2])
```

```
## [1] -0.137
```

```
ssh = (nh/(nh - 1))*(yes_ph*no_ph)
hat_var_ph = ssh/(nh-1)

B = 2*sqrt(hat_var_ph[1] + hat_var_ph[2])
B
```

```
## [1] 0.06487289
```

As we can see from above the difference between anesthesiologists from residents is, **-0.137**, with a bound of error **0.6487289**, and a confidence interval **(-0.07212711, -0.2018729)**. We can deem this as not significant.

Part C

```
(yes_ph[1] - yes_ph[3])
```

```
## [1] -0.095
```

```
ssh = (nh/(nh - 1))*(yes_ph*no_ph)
hat_var_ph = ssh/(nh-1)
B = 2*sqrt(hat_var_ph[1] + hat_var_ph[3])
B
```

```
## [1] 0.0356482
```

As we can see from above the difference between anesthesiologists from nurse anesthestists is, **-0.095**, with a bound of error **0.0356482**, and a confidence interval **(-0.1306482,-0.0593518)**. We can deem this as not significant.

Question 4

In the same survey discussed in Exercise 5.31, the respondents were asked for the longest continuous time (in hours) of administering anesthesia without a break over the last six months. A summary of the results is as follows:
a. Estimate the mean time for the population of those giving anesthesia, with an estimated bound on the error.
b. Do residents have a significantly higher average than the other groups? Justify your answer statistically.

Solution

Part A

```
nh = c(1347, 163, 1095)
n = sum(nh)
ah = nh/n
sd = c(0.15, 0.35, 0.11)

ybar_h = c(7.63, 7.74, 6.55)
nh = c(1347, 163, 1095)
ybarSt = sum(ah*ybar_h)
ybarSt
```

```
## [1] 7.18291
```

```
hat_var_ybarSt = sum((ah^2)*((sd^2)/nh))
hat_var_ybarSt
```

```
## [1] 9.361077e-06
```

```
B = 2*sqrt(hat_var_ybarSt)
B

ybarSt + B
ybarSt - B
```

The estimated mean time for the population of those giving anesthesia is **7.18291** with a bound on error of **0.006119175**. Furthermore we have a confidence interval of **(7.176791, 7.189029)**.

Part B

This is for comparing the significance between the resident and the anesthesiologist.

```
ybar_h[2] - ybar_h[1]
```

```
## [1] 0.11
```

```
B = 2*sqrt(sd[1]/nh[1] + sd[2]/nh[2])
B
```

```
## [1] 0.09504942
```

```
ybar_h[2] - ybar_h[1] - B
```

```
## [1] 0.01495058
```

```
ybar_h[2] - ybar_h[1] + B
```

```
## [1] 0.2050494
```

As shown above, we have a difference **0.11** with a bound on error of **0.09504942**, with a confidence interval, **(0.01495058, 0.2050494)**. As we can see we do see some difference hence we can say there is a significantly higher average.

This is for comparing the significance between the resident and the nurse anesthetist.

```
ybar_h[2] - ybar_h[3]
```

```
## [1] 1.19
```

```
B = 2*sqrt(sd[3]/nh[3] + sd[2]/nh[2])
B
```

```
## [1] 0.09481974
```

```
ybar_h[2] - ybar_h[3] - B
```

```
## [1] 1.09518
```

```
ybar_h[2] - ybar_h[3] + B
```

```
## [1] 1.28482
```

As shown above, we have a difference **1.19** with a bound on error of **0.09481974**, with a confidence interval, **(0.09481974, 1.28482)**. As we can see we do see some difference hence we can say there is a significantly higher average.