AGYEMANG ERIC

MAT 450 - FINAL EXAM

① ⓐ for the percentage of defective pieces in use, we have $\hat{\bar{y}}_r = \frac{\sum_{i\in S} t_i}{\sum_{i\in S} M_i} = \frac{80+210+40+120+10}{65+82+52+91+62}$

$= \frac{460}{3{,}520} = \boxed{0.13068} = \underline{\underline{13.068\%}}$

FACTORY	M_i	m_i	a_i	$\bar{y}_i$	$\hat{t}_i$
1	650	65	8	$8/65$	80
2	820	82	21	$21/82$	210
3	520	52	4	$4/52$	40
4	910	91	12	$12/91$	120
5	620	62	1	$1/62$	10
TOTALS	3520	352	46		

$M_i = 10\times m_i$ and $\bar{y}_i = \frac{a_i}{m_i}$ $\hat{t}_i = M_i\bar{y}_i$

ⓑ The 95% CI for $\hat{\bar{y}}_r$ above is given by

$\hat{\bar{y}}_r \pm 1.96\, SE(\hat{\bar{y}}_r)$

and $SE(\hat{\bar{y}}_r) = \sqrt{\hat{V}(\hat{\bar{y}}_r)}$ and we have the

$$\hat{V}(\hat{\bar{y}}_r) = \frac{1}{\bar{M}_s^2}\left[\left(1-\frac{n}{N}\right)\frac{s_r^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i\in S} M_i^2\left(1-\frac{m_i}{M_i}\right)\frac{s_i^2}{m_i}\right]$$

where $s_r^2 = \frac{1}{n-1}\sum(\hat{t}_i - M_i\hat{\bar{y}}_r)^2 = \frac{1}{(5-1)}\left[(80-650\times0.13068)^2 + (210-820\times0.13068)^2 + (40-520\times0.13068)^2\right] = 4406.90575$

Next we ignore the second part with $s_i^2 = \frac{1}{m_i - 1} \sum_{j \in S_i} (y_{ij} - \bar{y}_i)^2$ since we do not have y_{ij} in our question.

$$\Rightarrow \hat{V}(\hat{\bar{y}}_r) = \frac{1}{\bar{M}^2}\left[\left(1-\frac{n}{N}\right)\frac{s_r^2}{n}\right] = \frac{1}{\left(\frac{3520}{5}\right)^2}\left[1-\frac{5}{36}\right]\frac{4106.90575}{5}$$

$$= 0.0014271$$

$$\therefore SE(\hat{\bar{y}}_r) = \sqrt{0.0014271} = 0.037778.$$

$\therefore$ 95% CI is given by $0.13068 \pm 1.96(0.037778)$

$= [0.056607, 0.204693]$ this contains $\bar{y}_r$.

(c) For 95% CI based on Jackknife points, this is given by $\hat{B} \pm 1.96\, SE(\hat{B})$ where $\hat{B} = \hat{\theta} = \bar{y}_r = .1308$

and $SE(\hat{B}) = \sqrt{\hat{V}_{JK}(\hat{\bar{y}}_r)}$ where $\hat{V}_{JK}(\hat{\bar{y}}_r) = \frac{n-1}{n}\sum_{j \in S}(\hat{B}_{(j)} - \hat{B})$

where $\hat{B}_{(j)}$ is an estimator for θ. and $\hat{B}_{(j)} = \bar{y}_{(j)}/\bar{x}_{(j)}$.

Now we have $\hat{B}_{(1)} = \frac{21+4+12+1}{82+51+91+62} = 38/287$

$\hat{B}_{(2)} = \frac{8+4+12+1}{65+52+91+62} = 5/54$ $\quad$ $\hat{B}_{(3)} = \frac{8+21+12+1}{65+82+91+62} = 7/50$

$\hat{B}_{(4)} = \frac{8+21+4+1}{65+82+52+62} = \frac{34}{261}$ $\quad$ $\hat{B}_{(5)} = \frac{8+21+4+12}{65+82+52+91} = 9/58$

So we consider the table below:

Factory	$\hat{B}_{(j)}$	
1	38/287	
2	5/54	
3	7/50	
4	34/261	
5	9/58	

Now, $V_{JK}(\hat{B}) = \frac{4}{5}\left[\left(\frac{39}{287} - 0.13068\right)^2 + \left(\frac{5}{54} - 0.13068\right)^2 + \left(\frac{7}{50} - 0.13068\right)^2 + \left(\frac{34}{261} - 0.13068\right)^2 + \left(\frac{9}{58} - 0.13068\right)^2\right]$

$= 0.0017124$ and $SE(\hat{B}) = \sqrt{0.0017124}$

$= 0.041381481$

$\therefore$ 95% CI is given by $0.13068 \pm 1.96(0.041381481)$

$= (0.049575, 0.2117848)$

(d) Assum a_i's are SRS, then $\hat{p} = \bar{y}_{srs} = \frac{46}{352}$

so 95% CI is given by $\hat{p} \pm 1.96 SE(\hat{p})$

But $SE(\hat{p}) = \sqrt{\frac{p(1-p)}{\sum m_i}} = \sqrt{\frac{0.13068(1-0.13068)}{352}}$

$= 0.017964934$

$\Rightarrow 0.13068 \pm 1.96(0.017964934) \Rightarrow (0.09545, 0.16589)$

(e) Compairing the CI.s in b, c, and d above, we find that the CI in d is overlaped by that in b which is also overlaped by that of the Jackknife ~~in b~~ in C so the CI of the SRS is narrow compared with the others so we can say is the ~~better~~. followed by that of b and then followed by that in C.

(f) For the median consider the empirical probability mass function (epmf) $\hat{f}(y) = \frac{\sum_{i \in s, y_i = y} w_i}{\sum_{i \in s} w_i}$ with the cumulative distribution function

$w_{ij} = \frac{NM_i}{nm_i}$, $t_i = M\bar{y}_i$, $t = \sum_{i=1}^{n} t_i$ $\hat{F}(y) = \sum_{x \leq y} \hat{f}(y)$ and $w_i = m_i w_{ij}$

Factory	M_i	m_i	a_i	w_{ij}	w_i	$\hat{t}_i$	$\hat{f}(y)$
1	650	65	8	72	4680	80	0.1847
2	820	82	21	72	5904	210	0.23295
3	520	52	4	72	3744	40	0.14773
4	910	91	12	72	6552	120	0.2585
5	620	62	1	72	4464	10	0.1761
TOTALS	3520	352	46		25344		

For the median we arrange a_i in ascending order and find $\hat{F}(y)$

factory	a_i	$\hat{F}(y)$	$\hat{f}(y)$	mi
5	1	0.1761	0.1761	62
3	4	0.32383	0.14773	52
1	8	0.50853	0.1847	65
4	12	0.76703	0.2585	91
2	21	0.99998 $\simeq 1$	0.23295	82

Using the interpolation method,

The median $m = y_1 + \frac{q - \hat{F}(y_1)}{\hat{F}(y_2) - \hat{F}(y_1)} \cdot (y_2 - y_1)$

But $\hat{F}(4) = 0.32383$ and $\hat{F}(8) = 0.50853$

So $m = 4 + \frac{0.5 - 0.32383}{0.50853 - 0.32383} \times (8-4) = \boxed{7.815}$

Considering the population in broader perspective,

$m = 52 + \frac{0.5 - 0.32383}{0.50853 - 0.32383}(65 - 52) = 64.3996$

(9) The 95% CI for the median is given by:

$\hat{F}(a_{0.5}) \pm 1.96\sqrt{\hat{v}(\hat{F}(a_{0.5}))}$ where $\hat{v}(\hat{F}(a_{0.5})) = \frac{0.5(1-0.5)}{352}$

$= 0.000710227$

For the sample $1.96\sqrt{\hat{v}\hat{F}(\hat{a}_{0.5})} = 1.96\sqrt{0.000710227}$

$= 0.052234176$

The lower confidence bound is $\hat{F}^{-1}(0.5 - \hat{v}(\hat{F}(\hat{a}_{0.5})))$

and upper $\hat{F}^{-1}(0.5 + \hat{v}(\hat{F}(\hat{a}_{0.5})))$

$\Rightarrow \hat{F}^{-1}(0.5 - 0.052234) = 4 + \frac{0.447766 - 0.32383}{0.50853 - 0.32383}(8-4)$

$= 6.68405$

$\hat{F}^{-1}(0.5 + 0.052234) = 8 + \frac{0.552234 - 0.5085}{0.76703 - 0.50853}(12-8)$

$= 8.67627$

$= (6.68405, 8.67627)$ as the CI

(2) (a) Let $R = \{2, 3\}$ and $\hat{p} = \frac{x}{n}$.

Given that $\{2, 3\}$ are selected then for the given observations, the population mean is estimated as

$$\hat{p} = \hat{\bar{y}}_{\psi} = \frac{1}{n} \sum_{i \in R} \bar{y}_i \quad \text{where } \bar{y}_i = \frac{\sum_{j \in S} y_{ij}}{m_i}$$

Then $\bar{y}_2 = \frac{7+4}{2} = 5.5$ and $\bar{y}_3 = \frac{7+2+9+4}{4} = 5.5$

$\therefore \bar{y}_{\psi} = \frac{1}{2}(5.5 + 5.5) = 5.5$

For the proportion of observations with values less or equal to 5, we have

PSU	m_i	y_{ij}
2	4	4
3	8	2, 4

where $\frac{x}{n} = \hat{p}$

$$\Rightarrow \hat{p} = \left(\frac{4}{2} + \frac{6}{4}\right) \times \frac{1}{2} = \boxed{1.75}$$

For 95% of $\hat{p}$, $\hat{V}(\hat{p}) = \frac{1}{n} \cdot \frac{1}{n-1} \sum_{i \in R} (\bar{y}_i - \bar{y}_{\psi})^2$.

$$= \frac{1}{2} \times \left[(2 - 1.75)^2 + (1.5 - 1.75)^2\right] = 0.0625$$

using $Z_{0.025}$ test to find CI,

$$\Rightarrow \hat{p} \pm 1.96\, SE(\hat{p}) \Rightarrow 1.75 \pm 1.96\sqrt{0.0625}$$

$$\Rightarrow \underline{(1.26,\ 2.24)}.$$

using $t_{0.025, 4}$ gives a 95% CI of

$$1.75 \pm 2.776\sqrt{0.0625} = \underline{(1.056,\ 2.444)}$$

2 (b) The probability of selecting psus {2, 3} for the without replacement is given by

$P(2 \text{ chosen } 1^{st}) \cdot P(3 \text{ chosen } 2^{nd} / 2 \text{ chosen } 1^{st}) + P(3 \text{ chosen } 1^{st})$
$\times P(2 \text{ chosen } 2^{nd} / 3 \text{ chosen } 1^{st})$.

$\Rightarrow \psi_2 \left(\frac{\psi_3}{1-\psi_2}\right) + \left(\psi_3 \times \left(\frac{\psi_2}{1-\psi_3}\right)\right)$

$= \frac{4}{22} \times \left[\frac{8/22}{1-4/22}\right] + 8/22 \times \left[\frac{4/22}{1-8/22}\right] = \left(4/22 \times 4/9\right) + \left(8/22 \times 2/7\right)$

$= \boxed{0.1847}$

(c) The probability of selecting psu 3 WOR is given by:

$P([1,3]) + P([2,3]) + P([3,4]) + P([3,1]) + P([3,2]) +$
$P([4,3])$.

$= 5/22 \left\{\frac{8/22}{1-5/22}\right] + \left[4/22 \times \left(\frac{8/22}{1-4/22}\right)\right] + 8/22\left[\frac{5/22}{1-8/22}\right] +$

$8/22 \left\{\frac{5/22}{1-8/22}\right] + 8/22\left[\frac{4/22}{1-8/22}\right] + 5/22\left[\frac{8/22}{1-5/22}\right]$

$= 20/187 + 8/99 + 10/77 + 10/77 + 8/77 + 20/187$

$= \boxed{0.6583}$

2 (d) Using the SYG form for the variance.

$$\hat{V}_{SYG}(\hat{t}_{HT}) = \frac{1}{2}\sum_{i=1}\sum_{\substack{k=1 \\ k\neq i}}(\pi_i\pi_k - \pi_{ik})\left(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k}\right)^2.$$

but $\pi_{ik} = \psi_i\left[\frac{\psi_k}{1-\psi_i}\right] + \psi_k\left[\frac{\psi_i}{1-\psi_k}\right]$.

we have the table:

i \ k	1	2	3	4	TOTALS
1	—	0.103981	0.236822	0.133689	0.474492
2	0.103981	—	0.184704	0.103981	0.392666
3	0.23682	0.184704	—	0.236822	0.658346
4	0.13369	0.103981	0.236822		0.474493

$$\hat{t}_{HT} = \sum_{i\in s}\frac{t_i}{\pi_i} = \frac{25}{0.392666} + \frac{38}{0.658346} = 121.38774$$

$$\hat{\bar{y}}_{HT} = \frac{\hat{t}_{HT}}{22} = \frac{121.38774}{22} = 5.51762496$$

Now $\hat{V}_{SYG}(\hat{t}_{HT}) = \frac{1}{2}\sum_{i\in s}\sum_{\substack{k\in s \\ k\neq i}}(\pi_i\pi_k - \pi_{ik})\left(\frac{t_i}{\pi_i} - \frac{t_k}{\pi_k}\right)^2$

$= 14.132927$

So $\hat{SE}_{SYG}(\hat{\bar{y}}_{HT}) = \frac{1}{22}\sqrt{\hat{V}_{SYG}(\hat{t}_{HT})} = \frac{1}{22}\sqrt{14.132927}$

$= 0.170881$

$\therefore$ 95% CI is $5.5176 \pm 1.96(0.170881) = \underline{(5.1827, 5.8525)}$

2 (e) $\hat{V}_{WR}(\hat{t}_{HT}) = \frac{n}{n-1}\sum_{i\in S}\left(\frac{t_i}{\pi_i} - \frac{\hat{t}_{HT}}{n}\right)^2$

$$= \frac{2}{2-1}\left(\left[\frac{25}{0.392666} - \frac{121.38774}{2}\right]^2 + \left(\frac{38}{0.658346} - \frac{121.3877}{2}\right)^2\right)$$

$$= 35.17372$$

The 95% CI is $5.517624 \pm 1.96\sqrt{\frac{35.17372}{22}}$

$\Rightarrow$ $(4.9878, 6.0474)$

(f) From d and e above the CI for the with-replacement is larger in interval than that of the without replacement. This may show that there is a small variance in the without replacement leading to a small interval hence is considered the best than the with replacement Confidence interval.

For design effect;

$$deff(\text{plan, statistic}) = \frac{V(\text{estimator from Sampling plan})}{V(\text{estimator from an SRS with same number of observ})}$$

for the denominator, $\hat{V}(\bar{y}_{SRS}) = \frac{1}{n}\cdot\frac{1}{n-1}\sum_{i\in R}(y_i - \bar{y}_\varphi)^2$

$$= \frac{1}{2}\times\left[\left((6.25-5.5)^2 + (4.75-5.5)^2\right)\right] = 0.5625$$

So the deff for (WR, SRS) $= \frac{35.17372}{0.5625} = 62.531$

and that of WOR $= \frac{14.132927}{0.5625} = 25.12520$

The design effect for the with replacement is larger than that of the WOR sampling which shows that the WOR is the better compared with the WR sampling. So we need more observation in the WR sampling in order to be in same precision with the SRS and this is more that what we need from that of the WOR sampling.