IS3001

IS3001

Sampling Techniques

Group Project - Group 01

S14330 – U.A.B.M GUNASEKARA

S14429 – M.R.N FERNANDO

S14331 – B.V.R.R.V GUNATHUNGA

S14351 – A.M.D.S MAHAMITHAWA

S14304 – M.L.S.R ABHAYAWARDHANA

Contents

Introduction

The dataset we're working with, is about maternity and child health. It has several variables that can be used to describe children's health.

The variables in the dataset are, "District sector ID" which is a single variable created by combining the District ID and the Sector ID, *Grama Niladhari* Division ID, Birth Weight, Mother's Education, Child's Gender, Wealth Index Quintile, Ethnicity, Province, Sector, a Boolean variable that says if the child has low birth weight, Mother's Age, a Boolean variable that says if the mother has received *Thriposha,* number of clinical visits, number of months the mother spent pregnant, mother's height, and mother's BMI.

The dataset was analyzed using Simple Random Sampling, Stratified Sampling, and Cluster Sampling separately. "*rsampcal*" and "*ssampcal*" functions in R were used for determining samples.

In simple random sampling, the "*rsampcalc*" function was used to determine the sample size, and the obtained sample size was 893.

Stratified Sampling is based on dividing the population into various strata, and individuals are selected randomly from these strata to suit the sample size. (In these cases, the strata must be homogenous, collectively exhaustive, and mutually exclusive.) Here, "Mother's BMI" was used as the stratifying variable, and individuals were randomly selected from the BMI ranges proportionally to the sizes of strata, to suit the sampling size (893) determined in the random sampling method.

In cluster sampling, unlike in stratified sampling homogeneity is external but heterogeneity is internal within the clusters. In the two-stage sampling design the population is partitioned into groups, like cluster sampling, but in this design new samples are taken from each cluster sampled. And here, initially the population is divided into N clusters based on the variable "province", a sample of n clusters are selected from N and then individual elements are selected from these clusters randomly. All of these methods are explained in detail, in the next parts of the report.

Methodology

Sample Size calculation

Normally when doing a survey, this equation is used for sample size calculations.

$$n = \frac{n_0}{1+\frac{n_0}{N}} \quad ; \quad n_0 = \left(\frac{Z_{\frac{\alpha}{2}} S^2}{e}\right)^2$$

n = Sample Size

N = Population Size

$Z_{\frac{\alpha}{2}}$ = Z value of the significance level

S = Data Variation

e = Margin of error

Simple Random Sampling

For this project since R software is used for calculations, rsampcalc function that is included in sampler package is used for calculate the sample size. Code line is used as mention below.

n=rsampcalc(nrow(data), e=3, ci=95)

For simple random sampling we use this directly to obtain the sample size.

Stratified Random Sampling

But in the Stratified sample technique the population should be divided into strata. Then observations are selected from each stratum proportionally to the size of each stratum to obtain a sample size of n.

To select observations from each stratum to obtain a sample size of n=893, ssampcalc function that is included in sampler package is used. That code line is mentioned below.

ssampcalc(df =data, n = 893, strata =`Mother's BMI`)

Two-Stage Cluster Sampling

When considering the cluster sample technique, Summation of selected clusters populations is taken as the population size of the cluster sampling. That sample size is calculated using rsampcalc function that is included in sampler package as SRS. And the two-stage sample sizes of each cluster are calculated according to the proportions of each clusters populations by using ssampcalc function that is included in sampler package that used in the stratified random sample.

n=rsampcalc(nrow(one_cluster_samp1),e=3,ci=95)

two_cluster_samp_size1 = ssampcalc(df=one_cluster_samp1, n ,strata=`Province`)

Stratification Variable

In stratified sampling the population of N sampling units is divided into H "layers" or strata, with N_h sampling units in stratum h. The values of $N_1, N_2, \ldots, N_H$ should be known.

$N_1 + N_2 + \cdots + N_H = N$ where *N = Total number of units in the entire population*

In stratified random sampling an SRS is taken independently from each stratum, so that n_h observations are randomly selected from the N_h population units in stratum h. S_h is defined to be the set of n_h units in the SRS for stratum h. The total sample size is,

$$n = n_1 + n_2 + \cdots + n_H$$

To divide the population into H strata, a stratification variable should be selected. It's important that the strata should not overlap and the variance within the strata should be minimum.

To meet the given requirements, "Mother's BMI" was selected as the stratification variable which contains 4 strata as below:

1. BMI < 18.5
2. 18.5 < BMI < 24.9
3. 24.9 < BMI < 29.9
4. 30 <= BMI

Cluster Variable

Cluster sampling involves dividing the specific population of interest into geographically distinct groups or clusters, such as neighborhoods or families. In the two-stage sampling design the population is partitioned into groups, like cluster sampling, but in this design new samples are taken from each cluster sampled. The clusters are the first stage units to be sampled, called primary or first sampling units and denoted by SU1. The second-stage units are the elements of those clusters, called sub-units, secondary or second sampling units and denoted by SU2.

Because the information is readily available, to divide the population geographically, the variable "Province" was selected as the cluster variable in this analysis. Then a sample of provinces were selected and then a sample of observations was obtained from each cluster to make it a two-stage cluster sample.

Results

Simple Random Sampling

	Sample 1		Sample 2		Population		
	Estimated Population Mean	SE	Estimated Population Mean	SE	Mean	SE	Proportion
Birth weight of children	2906.1	15.539	2931.8	15.479	2912.705	6.4540	
Mother's age	31.085	0.2008	31.431	0.1955	30.9706	0.0794	
Number of clinical visits	8.0829	0.0653	8.1937	0.0672	8.02937	0.02658	
Number of months pregnant	8.9888	0.012	8.9922	0.0119	8.9919	0.005186	
	Estimated Proportion	SE	Estimated Proportion	SE			
Mother's Education							
• Higher	0.319148	0.0156	0.2989	0.0153			0.2915
• Never attended school	0.008958	0.0032	0.0067	0.0027			0.0066
• Pre school	0.001119	0.0011	-	-			0.0003
• Primary	0.039193	0.0065	0.0279	0.0055			0.0400
• Secondary	0.631578	0.0162	0.6662	0.0158			0.6614
Child's Gender							
• Female	0.47032	0.0167	0.4658	0.0167			0.4900
• Male	0.52968	0.0167	0.5341	0.0167			0.5099
Wealth Index Quintile							
• Fourth	0.18813	0.0131	0.2015	0.0134			0.1984
• Highest	0.18477	0.0130	0.1780	0.0128			0.1702
• Lowest	0.21725	0.0138	0.2239	0.0140			0.2363
• Middle	0.1959	0.0133	0.1937	0.0132			0.1913
• Second	0.21389	0.0137	0.2026	0.0135			0.2036
Ethnicity							
• Indian Tamil	0.04031	0.0066	0.0279	0.0055			0.0288
• Malay	0.0022	0.0016	0.0022	0.0016			0.0012
• Sinhala	0.67637	0.0157	0.6808	0.0156			0.6667
• Sri Lanka Muslim	0.11198	0.0106	0.0985	0.0100			0.1057
• Sri Lanka Tamil	0.16909	0.0126	0.1903	0.0131			0.1955
• Burger	-	-	-	-			
Province							
• Central	0.1887	0.0108	0.1209	0.0109			0.1221
• Eastern	0.1075	0.0104	0.0873	0.0095			0.1085
• North-central	0.0638	0.0082	0.0761	0.0089			0.0675

• North-western	0.1063	0.0103	0.1265	0.0111			0.1129
• Northern	0.0974	0.0099	0.1265	0.0111			0.1103
• Sabaragamuwa	0.0918	0.0097	0.0940	0.0098			0.0874
• Southern	0.1321	0.0113	0.1019	0.0101			0.1154
• Uva	0.0671	0.0084	0.0582	0.0078			0.0679
• Western	0.2150	0.0138	0.2082	0.0136			0.2076
Sector							
• Estate	0.0716	0.0086	0.0526	0.0075			0.0605
• Rural	0.7413	0.0147	0.7760	0.0140			0.7763
• Urban	0.1870	0.0131	0.1713	0.0126			0.1630
Low Birth Weight	0.1601	0.0123	0.1724	0.0126			0.1715
Mother's height							
• Average	0.5475	0.0167	0.5867	0.0165			0.5602
• Short	0.0582	0.0078	0.0526	0.0075			0.0694
• Tall	0.3941	0.0164	0.3605	0.0161			0.3703
Mother's BMI							
• <18.5	0.1243	0.0110	0.1220	0.0110			0.1173
• 18.5-24.9	0.4613	0.0167	0.4624	0.0167			0.4867
• 24.9-29.9	0.3113	0.0155	0.3101	0.0155			0.2886
• 30 and over	0.1030	0.0102	0.1052	0.0103			0.1072
	Estimated Totals	SE	Estimated Totals	SE	Estimated Totals	SE	
Low Birth Weight							
• Yes	143	10.965			934		
• No					4512		
Child's Gender							
• Female	420	14.924			2669		
• Male	473				2777		
Regression							
• Intercept	-76.78		221.3				
• No of months pregnant	331.84		301.4				
• Mean birth weight	2907.09		2931.46				

- Estimated mean value of birth weight for SRS 1 is 2906.1 g which is less than the population value. Also estimated mean value of birth weight for SRS 2 is 2931.8 g which is greater than the actual value. So, there is a difference between these two estimated values.
- Estimated proportions of SRS 1 for BMI levels are BMI < 18.5 - 0.1243, BMI 18.5-24.9 - 0.4613, BMI 24.9-29.9 - 0.3113 and BMI 30 and over - 0.1030. When compared with actual values, these estimated values are closer to them but not exact values. Also estimated proportions were obtained for SRS 2, those values also close to actual values as well as SRS 1.

- The estimated low birth weight proportions for SRS 1 and SRS 2, are 0.1601 and 0.1724 which has a small difference between them but both values are close to actual proportion.

Stratified Sampling

	Sample 1		Sample 2		Population		
	Estimated Population Mean	SE	Estimated Population Mean	SE	Mean	SE	Proportion
Birth weight of children	2931.2	15.624	2917.5	15.529	2912.705	6.4540	
Mother's age	30.917	0.1956	30.828	0.1916	30.9706	0.0794	
Number of clinical visits	7.9262	0.064	8.0738	0.0625	8.02937	0.02658	
Number of months pregnant	8.9787	0.0129	8.9787	0.0151	8.9919	0.005186	
	Estimated Proportion	SE	Estimated Proportion	SE			
Mother's education							
• Higher	0.2964	0.0153	0.2953	0.0152			0.2915
• Never attended school	0.0100	0.0033	0.0022	0.0016			0.0066
• Pre school	-	-	-				0.0003
• Primary	0.0503	0.0073	0.0302	0.0057			0.0400
• Secondary	0.6431	0.0160	0.6722	0.0157			0.6614
Child's gender							
• Female	0.4932	0.0167	0.4865	0.0167			0.4900
• Male	0.5067	0.0167	0.5134	0.0167			0.5099
Wealth index quintile							
• Fourth	0.1901	0.0131	0.2058	0.0135			0.1984
• Highest	0.1845	0.0129	0.1599	0.0122			0.1702
• Lowest	0.2315	0.0141	0.2203	0.0138			0.2363
• Middle	0.2058	0.0135	0.2058	0.0135			0.1913
• Second	0.1879	0.0130	0.2080	0.0135			0.2036
Ethnicity							
• Indian Tamil	0.0268	0.0054	0.0279	0.0055			0.0288
• Malay	-	-	0.0022	0.0016			0.0012
• Sinhala	0.6655	0.0158	0.6666	0.0157			0.6667
• Sri Lanka moor/Muslim	0.0973	0.0099	0.1118	0.0105			0.1057
• Sri Lanka Tamil	0.2069	0.0136	0.1912	0.0132			0.1955
• Burger	0.1901	0.0019	-	-			
Province							
• Central	0.1152	0.0107	0.1174	0.0108			0.1221
• Eastern	0.1118	0.0105	0.1163	0.0107			0.1085
• North-central	0.0637	0.0082	0.0671	0.0084			0.0675

• North-western	0.1196	0.0109	0.1029	0.0102			0.1129
• Northern	0.0995	0.0100	0.1040	0.0102			0.1103
• Sabaragamuwa	0.0872	0.0094	0.0783	0.0089			0.0874
• Southern	0.1241	0.0110	0.1219	0.0109			0.1154
• Uva	0.0872	0.0094	0.0715	0.0086			0.0679
• Western	0.1912	0.0132	0.2203	0.0139			0.2076
Sector							
• Estate	0.0659	0.0083	0.0604	0.0080			0.0605
• Rural	0.7639	0.0142	0.7751	0.0140			0.7763
• Urban	0.1700	0.0126	0.1644	0.0124			0.1630
Low birth weight	0.1554	0.012	0.1487	0.0119			0.1715
Mother's height							
• Average	0.5626	0.0166	0.5693	0.0166			0.5602
• Short	0.0648	0.0082	0.0637	0.0082			0.0694
• Tall	0.3724	0.0162	0.3668	0.0161			0.3703
Mother's BMI							
• <18.5	0.1174	0	0.1174	0			0.1173
• 18.5-24.9	0.4866	0	0.4865	0			0.4867
• 24.9-29.9	0.2885	0	0.2885	0			0.2886
• 30 and over	0.1073	0	0.1073	0			0.1072
	Estimated Totals	SE	Estimated Totals	SE	Estimated Totals	SE	
Low Birth Weight							
• Yes	845	65.477	808	64.75	934		
• No	4601		4638		4512		
Child's Gender							
• Female	2681	90.88	2644	90.76	2669		
• Male	2765	90.88	2802	90.76	2777		
Regression							
• Intercept	-102.4		326.4				
• No of months pregnant	337.9		288.6				
• Mean birth weight	2935.97		2921.46				

- Estimated mean value of birth weight for Stratified sample 1 is 2931.2 g. When compared with actual mean value, this value is a little bit higher. Also estimated mean value of birth weight for Stratified sample 2 is 2917.5 g also greater than actual value. So, there is a difference between these two estimated values.
- Estimated low birth proportions were obtained as 0.1554 and 0.1487 for stratified samples 1 and sample 2 respectively. When compared with actual proportion, these estimates are low values and also have a difference but close to actual value.

Two-Stage Cluster Sampling

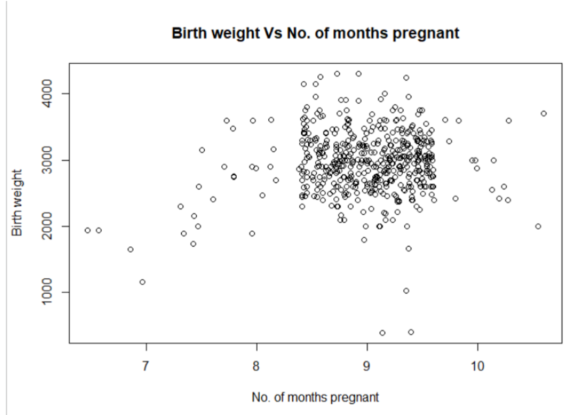
	Sample 1		Sample 2		Population		
	Estimated Population Mean	SE	Estimated Population Mean	SE	Mean	SE	Proportion
Birth weight of children	2881.5	29.462	2874	15.72	2912.705	6.4540	
Mother's age	30.663	0.2442	30.936	0.194	30.9706	0.0794	
Number of clinical visits	7.9123	0.1894	8.0913	0.095	8.02937	0.02658	
Number of months pregnant	8.946	0.0294	8.9531	0.0224	8.9919	0.005186	
	Estimated Proportion	SE	Estimated Proportion	SE			
Mother's education							
• Higher	0.2537	0.0156	0.2555	0.0164			0.2915
• Never attended school	0.0107	0.0061	0.0091	0.0037			0.0066
• Pre school	0.0013	0.0013	-	-			0.0003
• Primary	0.0485	0.0172	0.0612	0.0116			0.0400
• Secondary	0.6855	0.02	0.6740	0.0126			0.6614
Child's gender							
• Female	0.4844	0.0311	0.5097	0.0122			0.4900
• Male	0.5155	0.0311	0.4902	0.0122			0.5099
Wealth index quintile							
• Fourth	0.1781	0.022	0.1825	0.0259			0.1984
• Highest	0.1120	0.0081	0.1186	0.0165			0.1702
• Lowest	0.2213	0.0372	0.2777	0.0304			0.2363
• Middle	0.2267	0.0249	0.1916	0.0124			0.1913
• Second	0.2618	0.0175	0.2294	0.0105			0.2036
Ethnicity							
• Indian Tamil	0.0229	0.0151	0.0599	0.0326			0.0288
• Malay	-	-	-	-			0.0012
• Sinhala	0.6990	0.1268	0.6453	0.1397			0.6667
• Sri Lanka moor/Muslim	0.1322	0.0502	0.1134	0.0542			0.1057
• Sri Lanka Tamil	0.1430	0.0889	0.1799	0.0953			0.1955
• Burger	0.0026	0.0017	0.0013	0.0014			
Province							
• Central	-	-	0.2451	0.2322			0.1221
• Eastern	0.2442	0.2319	0.2177	0.2139			0.1085
• North-central	0.1525	0.1621	-	0.2203			0.0675
• North-western	0.2537	0.2377	0.2268	0.1810			0.1129
• Northern	-	-	-	-			0.1103
• Sabaragamuwa	0.1970	0.1989	0.1747	0.1810			0.0874

• Southern	-	-	-	-			0.1154
• Uva	0.1039	0.1621	0.1355	0.1467			0.0679
• Western	-	-	-				0.2076
Sector							
• Estate	0.0472	0.0296	0.1160	0.0697			0.0605
• Rural	0.8488	0.0487	0.7783	0.0693			0.7763
• Urban	0.1039	0.0545	0.1056	0.0514			0.1630
Low birth weight	0.1794	0.0117	0.2007	0.0087			0.1715
Mother's height							
• Average	0.5587	0.0321	0.5723	0.0097			0.5602
• Short	0.0728	0.0127	0.0834	0.0104			0.0694
• Tall	0.3684	0.0365	0.3441	0.0198			0.3703
Mother's BMI							
• <18.5	0.1376	0.0168	0.1186	0.0135			0.1173
• 18.5-24.9	0.5101	0.008	0.4837	0.0080			0.4867
• 24.9-29.9	0.2672	0.0138	0.3011	0.0140			0.2886
• 30 and over	0.0850	0.0121	0.0964	0.0144			0.1072
	Estimated Totals	SE	Estimated Totals	SE	Estimated Totals	SE	
Low Birth Weight							
• Yes	783	84.64	985	130.14	934		
• No	4663				4512		
Child's Gender							
• Female	2114	260.16	2502	224.32	2669		
• Male	3332	284.55	2944	273.30	2777		
<u>Regression</u>							
• Intercept	-110.8		-27.75				
• No of months pregnant	334.5		324.11				
• Mean birth weight	2896.99		2886.62				

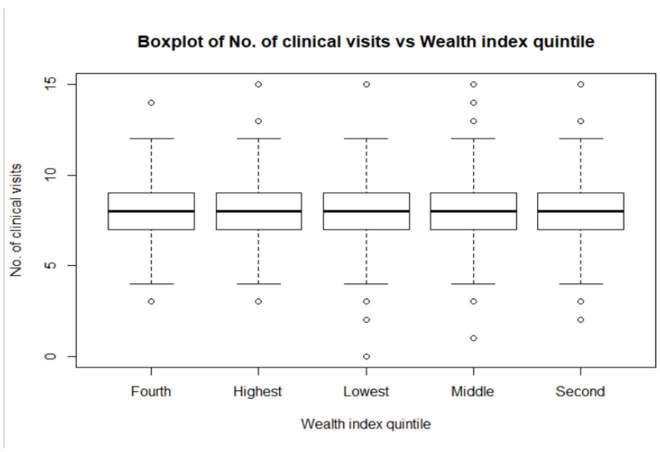
- According to obtained results for cluster sampling, the estimated mean of birth weight of cluster sample 1 is 2881.5 g. This is a low value compared to the actual value. Also, the estimated mean of birth weight of cluster sample 2 is 2874 g also a low value compared to the actual value. There is a difference between these two estimated values.
- Estimated proportions of BMI levels for cluster sample 1, are BMI < 18.5 - 0.1376, BMI 18.5-24.9 - 0.5101, BMI 24.9-29.9 - 0.2672, BMI 30 and over - 0.0850. When compared with actual proportions, these values are close to them. Also estimated proportions of BMI for cluster sample 2, are closer to population proportions of BMI as well as estimated proportions of cluster sample 1.

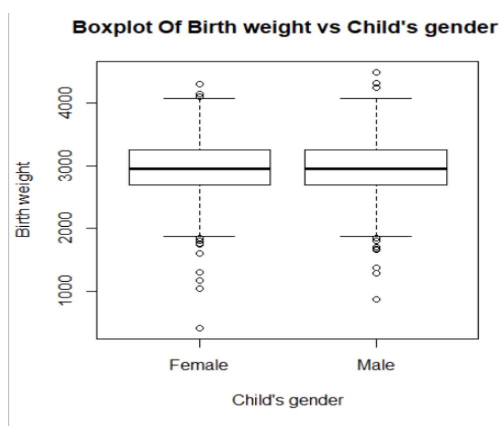
- Estimated proportions for low birth weights are 0.1794 and 0.2007 for two cluster samples. Estimate of sample 1 is close to the actual value but the estimate of sample 2 is little bit greater than the actual value, and also has a difference between two estimates.

Graphical Analysis - Stratified Sampling



The graph of Birth weight vs. No of months pregnant shows a slightly positive relationship, suggesting that there's a chance of getting a higher birth weight when the no. of months pregnant is higher. The graph also shows a range in both variables where majority of the results fall into.

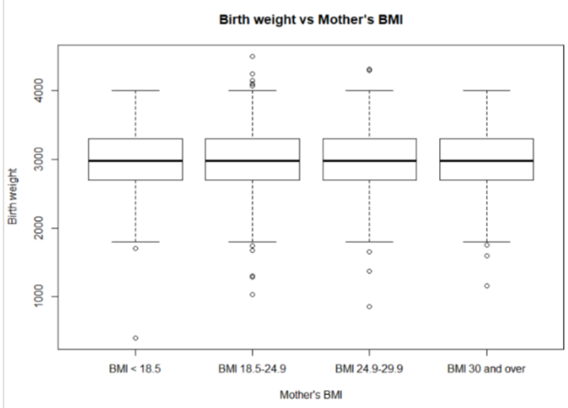


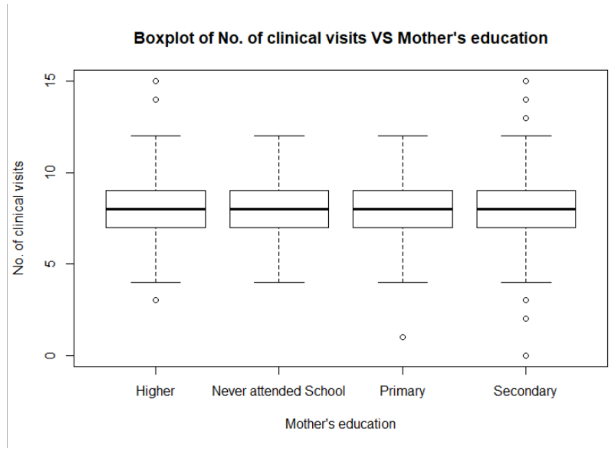


In the boxplot of no. Of clinical visits vs. Wealth index suggests that all the wealth index quartile groups have the same centers and the same spreads. The same thing goes for the boxplot of birthweight vs. Child's gender, the boxplot of birthweight vs. Mother's BMI, and the boxplot of clinical visits vs. Mother's education. All apparently having the same centers and spreads. This might also be happening because of the magnitude of the y-axis scales. Nevertheless, the outliers are also included here and some little variations can be seen from the outliers.

Considering the outliers, in the boxplot of clinical visits vs. Wealth index quartile, a little decrease can be seen in the "lowest" group in the no. of clinical visits and a little increase in the "middle" group. These are still just outliers and in the current y-axis scale, the boxes still show the same center and spread.

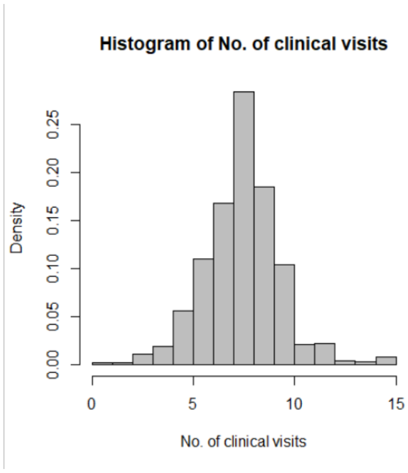
Considering the outliers in the boxplot of birthweight vs. Child's gender, a little decrease in birthweight can be seen in the female group compared to the male group.





Again, in the boxplot of clinical visits vs. Mother's education, a little increase in the number of clinical visits can be seen in the “higher” group and a little more spread can be seen in the “secondary” group. But again, these are interpreted by just observing the outliers which might not be a very good source, but given the fact that the boxes are identical in all the groups, observing the outliers seems adequate.

In the same way, in the boxplot of birthweight vs. Mother's BMI, the groups (24.9-29.9) and (30 and over) show a little decrease in birthweight, while the group (18.5-24.9) shows a little more spread.



The histogram of no. of clinical visits, shows an almost perfect normal distribution, suggesting that there’s a higher number of observations for 7-8 clinical visits.

The histogram of birth weight shows a negatively skewed distribution. Which suggests that most of the children are born with a weight around 2500-3000g.

Conclusion

On the basis of these results, several conclusions can be made on this analysis.

- Mean birth weight of children is 2931.2g.
- Average age of a pregnant mother is 31 years.
- On average 8 number of clinical visits are attended by a mother.
- Average duration of pregnancy is 9 months.
- It can be concluded that premature births may result low weight births.
- There is no significant difference in the birth weights of male children and female children.
- Only the mothers who have completed secondary or higher education have attended more than 12 clinical visits. This may suggest that higher the education level, higher the influence to attend clinical visits.
- Variability of birth weights of children of mothers in the BMI range 18.5-24.9 is higher than the other BMI ranges.
- Total number of female child births is relatively low compared to the male child births.
- Proportion of mothers from rural areas is significantly higher than the proportions of urban and estate.
- Proportion of mothers who have completed secondary or higher education is more than 0.9.
- Results of the analysis does not differ significantly with the method of sampling. But the standard error of the estimations in the two-staged cluster sampling is higher when compared to the other two which should be considered when conducting the analysis.
- It can be concluded that any of the three methods used in this analysis can be used to carry out an efficient analysis with subject to the constraints occur while conducting the analysis such as cost, effort and time.

R code

SRS

```
attach(Dataset_01)
data=Dataset_01

#----------------------Population Statistics----------------------------------#
N=nrow(data)
mean(data$`Birth weight`,na.rm=TRUE)
sqrt(var(data$`Birth weight`,na.rm=TRUE)/N)
mean(data$`Mother's age`,na.rm=TRUE)
sqrt(var(data$`Mother's age`,na.rm=TRUE)/N)
mean(data$`No. of clinical visits`,na.rm=TRUE)
sqrt(var(data$`No. of clinical visits`,na.rm=TRUE)/N)
mean(data$`No. of months pregnant`,na.rm=TRUE)
sqrt(var(data$`No. of months pregnant`,na.rm=TRUE)/N)

data.frame((table(data$`Mother's education`))/length(data$`Mother's education`))
data.frame((table(data$`Child's gender`))/length(data$`Child's gender`))
data.frame(table(data$`Wealth index quintile`)/length(data$`Wealth index quintile`))
data.frame(table(data$Ethnicity)/length(data$Ethnicity))
data.frame(table(data$Province)/length(data$Province))
data.frame(table(data$Sector)/length(data$Sector))
data.frame(table(data$`Low birth weight - 1 if it is a low birth weight, 0 otherwise`)/length(data$`Low birth weight - 1 if it is a
low birth weight, 0 otherwise`))
data.frame(table(data$`Mother's height`)/length(data$`Mother's height`))
data.frame(table(data$`Mother's BMI`)/length(data$`Mother's BMI`))

#-----------Totals
table(`Low birth weight - 1 if it is a low birth weight, 0 otherwise`)
table(`Child's gender`)

library(sampler)
#------------------------------------SRS---------------------------------------#
#----SAMPLE1-------
n=rsampcalc(nrow(data),e=3,ci=95)
set.seed(013001)
SRS_Sample1 = rsamp(df = data,n, rep = FALSE)
SRS_Sample1
library(survey)
SRS1=svydesign(id=~1,data=SRS_Sample1)
svymean(~`Birth weight`,design=SRS1,na.rm=TRUE)
svymean(~`Mother's age`,design=SRS1,na.rm=TRUE)
svymean(~`No. of clinical visits`,design=SRS1,na.rm=TRUE)
svymean(~`No. of months pregnant`,design=SRS1,na.rm=TRUE)
svymean(~`Mother's education`,design=SRS1,na.rm=TRUE)
svymean(~`Child's gender`,design=SRS1,na.rm=TRUE)
svymean(~`Wealth index quintile`,design=SRS1,na.rm=TRUE)
svymean(~`Ethnicity`,design=SRS1,na.rm=TRUE)
svymean(~`Province`,design=SRS1,na.rm=TRUE)
```

```
svymean(~`Sector`,design=SRS1,na.rm=TRUE)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=SRS1,na.rm=TRUE)
svymean(~`Mother's height`,design=SRS1,na.rm=TRUE)
svymean(~`Mother's BMI`,design=SRS1,na.rm=TRUE)
#-----------Totals
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=SRS1, na.rm=FALSE)
svytotal(~`Child's gender`, design=SRS1)

#Regression Estimation
SRS_Sample1
plot(SRS_Sample1$`No. of clinical visits`,SRS_Sample1$`Birth weight`)
#plot(SRS_Sample$`Birth weight`,SRS_Sample$`Wealth index quintile`)
plot(SRS_Sample1$`Mother's age`,SRS_Sample1$`Birth weight`)
plot(SRS_Sample1$`No. of months pregnant`,SRS_Sample1$`Birth weight`)
SLR1 = lm(`Birth weight`~`No. of months pregnant`,SRS_Sample1)
SLR1
mean_BirthWeight=-76.78+331.84*mean(data$`No. of months pregnant`)
mean_BirthWeight

#------SAMPLE2-------
n=rsampcalc(nrow(data),e=3,ci=95)
set.seed(023001)
SRS_Sample2 = rsamp(df = data,n, rep = TRUE)
SRS_Sample2

SRS2=svydesign(id=~1,data=SRS_Sample2)

svymean(~`Birth weight`,design=SRS2,na.rm=TRUE)
svymean(~`Mother's age`,design=SRS2,na.rm=TRUE)
svymean(~`No. of clinical visits`,design=SRS2,na.rm=TRUE)
svymean(~`No. of months pregnant`,design=SRS2,na.rm=TRUE)
svymean(~`Mother's education`,design=SRS2,na.rm=TRUE)
svymean(~`Child's gender`,design=SRS2,na.rm=TRUE)
svymean(~`Wealth index quintile`,design=SRS2,na.rm=TRUE)
svymean(~`Ethnicity`,design=SRS2,na.rm=TRUE)
svymean(~`Province`,design=SRS2,na.rm=TRUE)
svymean(~`Sector`,design=SRS2,na.rm=TRUE)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=SRS2,na.rm=TRUE)
svymean(~`Mother's height`,design=SRS2,na.rm=TRUE)
svymean(~`Mother's BMI`,design=SRS2,na.rm=TRUE)
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=SRS2, na.rm=TRUE)
svytotal(~`Child's gender`, design=SRS1)
#Regression Estimation
SRS_Sample2
plot(SRS_Sample2$`No. of clinical visits`,SRS_Sample2$`Birth weight`)
#plot(SRS_Sample2$`Birth weight`,SRS_Sample2$`Wealth index quintile`)
plot(SRS_Sample2$`Mother's age`,SRS_Sample2$`Birth weight`)
plot(SRS_Sample2$`No. of months pregnant`,SRS_Sample2$`Birth weight`)

SLR2 = lm(`Birth weight`~`No. of months pregnant`,SRS_Sample2)
SLR2
mean_BirthWeight=221.3+301.4*mean(data$`No. of months pregnant`)
mean_BirthWeight
barplot(SRS1)
```

```
library(rmeta)
forestplot(bw1,coef(bw1),coef(bw1)-sqrt(2)*SE(bw1),coef(bw1)+sqrt(2)*SE(bw1),align=c("1","1"),zero=500)
```

Stratified

```
attach(Dataset_01)
data=Dataset_01
library(sampler)
#--------------------------------Stratified------------------------------------#
#---------SAMPLE1
Strat_size = ssampcalc(df =data,n = 893, strata =`Mother's BMI`)
Strat_size
set.seed(013001)
stratified1 = ssamp(df=data,n=893, strata =`Mother's BMI`)
stratified1
pw1=0
for(i in 1:893){
   if(stratified1$`Mother's BMI`[i]=="BMI < 18.5"){
   pw1 = round((639/105),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 18.5-24.9"){
   pw1 = round((2651/435),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 24.9-29.9"){
   pw1 = round((1572/258),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 30 and over"){
   pw1 = round((584/96),2)
  }
  #pw=c(print(pw))
}

strat1 = cbind(stratified1,pw1)
strat1


library(survey)
sample1=svydesign(id=~1,strata =~`Mother's BMI`,weights=~`pw1` ,data=strat1)
svymean(~`Birth weight`,design=sample1)
svymean(~`Mother's age`,design=sample1)
svymean(~`No. of clinical visits`,design=sample1)
svymean(~`No. of months pregnant`,design=sample1)
svymean(~`Mother's education`,design=sample1)
svymean(~`Child's gender`,design=sample1)
svymean(~`Wealth index quintile`,design=sample1)
svymean(~`Ethnicity`,design=sample1)
```

```
svymean(~`Province`,design=sample1)
svymean(~`Sector`,design=sample1)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=sample1)
svymean(~`Mother's height`,design=sample1)
svymean(~`Mother's BMI`,design=sample1)
#-----------Totals
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=sample1)
svytotal(~`Child's gender`, design=sample1)


#Regression Estimation
sample1
plot(stratified1$`No. of clinical visits`,stratified1$`Birth weight`)
#plot(stratified1$`Birth weight`,stratified1$`Wealth index quintile`)
plot(stratified1$`Mother's age`,stratified1$`Birth weight`)

plot(stratified1$`No. of months pregnant`,stratified1$`Birth weight`)
SLR1 = lm(`Birth weight`~`No. of months pregnant`,stratified1)
SLR1

mean_BirthWeight=-102.4+337.9*mean(data$`No. of months pregnant`)
mean_BirthWeight

#SAMPLE2
Strat_size = ssampcalc(df =data,n = 893, strata =`Mother's BMI`)
Strat_size
set.seed(023001)
stratified2 = ssamp(df=data,n=893, strata =`Mother's BMI`)
stratified2
pw2=0
for(i in 1:893){
  if(stratified1$`Mother's BMI`[i]=="BMI < 18.5"){
    pw2 = round((639/105),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 18.5-24.9"){
    pw2 = round((2651/435),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 24.9-29.9"){
    pw2 = round((1572/258),2)
  }else if(stratified1$`Mother's BMI`[i]=="BMI 30 and over"){
    pw2 = round((584/96),2)
  }
  pw2=c(print(pw2))
}

strat2 = cbind(stratified2,pw2)
```

```
strat2
#------Sample2
sample2=svydesign(id=~1,strata =~`Mother's BMI`, weights=~`pw2`, data=strat2 )

svymean(~`Birth weight`,design=sample2)
svymean(~`Mother's age`,design=sample2)
svymean(~`No. of clinical visits`,design=sample2)
svymean(~`No. of months pregnant`,design=sample2)
svymean(~`Mother's education`,design=sample2)
svymean(~`Child's gender`,design=sample2)
svymean(~`Wealth index quintile`,design=sample2)
svymean(~`Ethnicity`,design=sample2)
svymean(~`Province`,design=sample2)
svymean(~`Sector`,design=sample2)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=sample2)
svymean(~`Mother's height`,design=sample2)
svymean(~`Mother's BMI`,design=sample2)
#===========Totals
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=sample2)
svytotal(~`Child's gender`, desig=sample2)

#------Regression Estimation
stratified2
plot(stratified2$`No. of clinical visits`,stratified2$`Birth weight`)
#plot(stratified2$`Birth weight`,stratified2$`Wealth index quintile`)
plot(stratified2$`Mother's age`,stratified2$`Birth weight`)
plot(stratified2$`No. of months pregnant`,stratified2$`Birth weight`)
SLR2 = lm(`Birth weight`~`No. of months pregnant`,stratified2)
SLR2

mean_BirthWeight=326.4+288.6*mean(data$`No. of months pregnant`)
mean_BirthWeight

#---------------------------Plots--------------------
#---------Boxplots

svyplot(strat1$`Birth weight`~strat1$`No. of months pregnant`, design=sample1, style="subsample" ,
main= "Birth weight Vs No. of months pregnant", ylab ="Birth weight",xlab="No. of months pregnant")

svyboxplot(`No. of clinical visits`~factor(`Wealth index quintile`), sample1, all.outliers=TRUE,
xlab="Wealth index quintile", ylab="No. of clinical visits", main="Boxplot of No. of clinical visits vs
Wealth index quintile")
```

```
svyboxplot(`Birth weight`~factor(`Child's gender`), sample1, all.outliers=TRUE, xlab="Child's gender",
ylab="Birth weight", main="Boxplot Of Birth weight vs Child's gender ")

svyboxplot(`Birth weight`~factor(`Mother's BMI`), sample1, all.outliers=TRUE, xlab="Mother's BMI",
ylab="Birth weight", main="Birth weight vs Mother's BMI")

svyboxplot(`No. of clinical visits`~factor(`Mother's education`), sample1, all.outliers=TRUE,
xlab="Mother's education", ylab="No. of clinical visits", main="Boxplot of No. of clinical visits VS
Mother's education")

#----------Histogram

svyhist(~`Birth weight`,sample1, xlab = "Birth Weight", main = "Histogram of Birth Weight", col = "grey")
#lines(svysmooth(~`Birth weight`, sample1, bandwidth=), lwd = 1)
svyhist(~`No. of clinical visits`,sample1, xlab = "No. of clinical visits", main = "Histogram of No. of clinical
visits", col = "grey")
```

Two stage cluster

```
attach(Dataset_01)
data=Dataset_01
#-------------------------Two Stage Cluster----------------------------------#
#-------------SAMPLE1
set.seed(013001)
clusters1 = sample(unique(data$Province),size=5,replace=F)
clusters1
one_cluster_samp1 = data[data$Province %in% clusters1,]
one_cluster_samp1
table(one_cluster_samp1$Province)
#n = nrow(one_cluster_samp1)
n=rsampcalc(nrow(one_cluster_samp1),e=3,ci=95)
library(sampler)
two_cluster_samp_size1 = ssampcalc(df=one_cluster_samp1,n ,strata=`Province`)
two_cluster_samp_size1
set.seed(013001)
two_cluster_samp1 = ssamp(df=one_cluster_samp1 ,n ,strata=`Province`)
two_cluster_samp1

pw1=0
for(i in 1:741){
  if(two_cluster_samp1$Province[i]=="Eastern"){
    pw1=round((591/181)*(9/5),2)
  }else if(two_cluster_samp1$Province[i]=="North-central"){
    pw1=round((368/113)*(9/5),2)
  }else if(two_cluster_samp1$Province[i]=="North-western"){
    pw1=round((615/188)*(9/5),2)
  }else if(two_cluster_samp1$Province[i]=="Sabaragamuwa"){
```

```
    pw1=round((476/146)*(9/5),2)
  }else if(two_cluster_samp1$Province[i]=="Uva"){
    pw1=round((370/113)*(9/5),2)
  }
  pw1 = c(pw1)
}

two_clust1 = cbind(two_cluster_samp1,pw1)
two_clust1

library(survey)
clust1=svydesign(id=~`Province`+`Grama Niladhari Division ID`, weights=~`pw1`, data=two_clust1)

svymean(~`Birth weight`,design=clust1)
svymean(~`Mother's age`,design=clust1)
svymean(~`No. of clinical visits`,design=clust1)
svymean(~`No. of months pregnant`,design=clust1)
svymean(~`Mother's education`,design=clust1)
svymean(~`Child's gender`,design=clust1)
svymean(~`Wealth index quintile`,design=clust1)
svymean(~`Ethnicity`,design=clust1)
svymean(~`Province`,design=clust1)
svymean(~`Sector`,design=clust1)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=clust1)
svymean(~`Mother's height`,design=clust1)
svymean(~`Mother's BMI`,design=clust1)
#-----------Totals
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=clust1)
svytotal(~`Child's gender`, desig=clust1)

#Regression Estimation
two_clust1
plot(two_clust1$`No. of clinical visits`,two_clust1$`Birth weight`)
#plot(two_clust1$`Birth weight`,two_clust1$`Wealth index quintile`)
plot(two_clust1$`Mother's age`,two_clust1$`Birth weight`)
plot(two_clust1$`No. of months pregnant`,two_clust1$`Birth weight`)
SLR1 = lm(`Birth weight`~`No. of months pregnant`,two_clust1)
SLR1

mean_BirthWeight=-110.8+334.5*mean(data$`No. of months pregnant`)
mean_BirthWeight

#-----------------SAMPLE2

set.seed(023001)
clusters2 = sample(unique(data$Province),size=5,replace=F)
clusters2
one_cluster_samp2 = data[data$Province %in% clusters2,]
one_cluster_samp2
```

```
table(one_cluster_samp2$Province)
#n = nrow(one_cluster_samp2)
n=rsampcalc(nrow(one_cluster_samp2),e=3,ci=95)

two_cluster_samp_size2 = ssampcalc(df=one_cluster_samp2, n ,strata=`Province`)
two_cluster_samp_size2

set.seed(023001)
two_cluster_samp2 = ssamp(df=one_cluster_samp2 , n ,strata=`Province`)
two_cluster_samp2

pw2=0
for(i in 1:767){
  if(two_cluster_samp2$Province[i]=="Eastern"){
    pw2=round(((591/167)*(9/5)),2)
  }else if(two_cluster_samp2$Province[i]=="Central"){
    pw2=round(((665/188)*(9/5)),2)
  }else if(two_cluster_samp2$Province[i]=="North-western"){
    pw2=round(((615/174)*(9/5)),2)
  }else if(two_cluster_samp2$Province[i]=="Sabaragamuwa"){
    pw2=round(((476/134)*(9/5)),2)
  }else if(two_cluster_samp2$Province[i]=="Uva"){
    pw2=round(((370/104)*(9/5)),2)
  }
  pw2 =(c(pw2))
}
two_clust2 = cbind(two_cluster_samp2,pw2)
two_clust2

library(survey)
clust2=svydesign(id=~`Province`+`Grama Niladhari Division ID`, weights=~`pw2`, data=two_clust2)

svymean(~`Birth weight`,design=clust2)
svymean(~`Mother's age`,design=clust2)
svymean(~`No. of clinical visits`,design=clust2)
svymean(~`No. of months pregnant`,design=clust2)
svymean(~`Mother's education`,design=clust2)
svymean(~`Child's gender`,design=clust2)
svymean(~`Wealth index quintile`,design=clust2)
svymean(~`Ethnicity`,design=clust2)
svymean(~`Province`,design=clust2)
svymean(~`Sector`,design=clust2)
svymean(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`,design=clust2)
svymean(~`Mother's height`,design=clust2)
svymean(~`Mother's BMI`,design=clust2)
#-------------Totals
svytotal(~`Low birth weight - 1 if it is a low birth weight, 0 otherwise`, design=clust2)
svytotal(~`Child's gender`, desig=clust2)
```

```
#-----------Regression Estimation
two_clust2
plot(two_clust2$`No. of clinical visits`,two_clust2$`Birth weight`)
#plot(two_clust2$`Birth weight`,two_clust2$`Wealth index quintile`)
plot(two_clust2$`Mother's age`,two_clust2$`Birth weight`)
plot(two_clust2$`No. of months pregnant`,two_clust2$`Birth weight`)
SLR2 = lm(`Birth weight`~`No. of months pregnant`,two_clust2)
SLR2

mean_BirthWeight=-27.75+324.11*mean(data$`No. of months pregnant`)
mean_BirthWeight
```