|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | |  | | --- | | **Appendix C**   1. **SAS CODES FOR DESCRIPTIVE STATISTICS OF THE DATA**   /\* Generated Code (IMPORT) \*/  /\* Source File: MAIN.xlsx \*/  /\* Source Path: /home/u45193287 \*/  /\* Code generated on: 4/21/20, 2:33 AM \*/  %web\_drop\_table(WORK.IMPORT1);  FILENAME REFFILE '/home/u45193287/MAIN.xlsx';  PROC IMPORT DATAFILE=REFFILE  DBMS=XLSX  OUT=WORK.IMPORT1;  GETNAMES=YES;  RUN;  PROC CONTENTS DATA=WORK.IMPORT1; RUN;  %web\_open\_table(WORK.IMPORT1);  \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  TITLE"Descriptive Statistics Of Main Dataset on City\_Mpg Based on Make";  proc means Data=WORK.IMPORT1 MEAN VAR STD CLM;  CLASS Make;  var City\_Mpg;  Run;  FILENAME REFFILE '/home/u45193287/HONDA.xlsx';  PROC IMPORT DATAFILE=REFFILE  DBMS=XLSX  OUT=WORK.IMPORT;  GETNAMES=YES;  RUN;  PROC CONTENTS DATA=WORK.IMPORT; RUN;  %web\_open\_table(WORK.IMPORT);  TITLE"Descriptive Statistics For HONDA on City\_Mpg Based on Class";  Proc means Data=WORK.IMPORT;  RUN;  proc means Data=WORK.IMPORT Median MEAN VAR STD CLM;  CLASS Class;  var City\_mpg;  run;  proc means Data=WORK.IMPORT Median MEAN VAR STD CLM;  CLASS Class;  var High\_mpg;  Run;  FILENAME REFFILE '/home/u45193287/TOYOTA.xlsx';  PROC IMPORT DATAFILE=REFFILE  DBMS=XLSX  OUT=WORK.IMPORT;  GETNAMES=YES;  RUN;  PROC CONTENTS DATA=WORK.IMPORT; RUN;  %web\_open\_table(WORK.IMPORT);  TITLE"Descriptive Statistics For TOYOTA on City\_mpg Based on Class";  proc means Data=WORK.IMPORT MEAN VAR STD CLM;  CLASS Class;  var City\_mpg;  run;  **# R-CODES FOR KERNEL DENSITY ESTIMATION**  **#The Kernel Density Estimation of Honda City Mpg**  den\_H1 <- density(HONDA$City\_mpg, bw = 4, na.rm = TRUE)  plot1=plot(den\_H1, main="Kernel Density of HONDA City\_mpg", ylab = "Density")  **#The Kernel Density Estimation of Honda Highway Mpg**  den\_H2 <- density(HONDA$High\_mpg, bw = 4, na.rm = TRUE)  plot1=plot(den\_H2, main="Kernel Density of HONDA Highway\_mpg", ylab = "Density")  #######################################################################  **#The Kernel Density Estimation of Toyota City Mpg**  den\_T1 <- density(TOYOTA$City\_mpg, bw = 4, na.rm = TRUE)  plot1=plot(den\_T1, main="Kernel Density of TOYOTA City\_mpg", ylab = "Density")  **#The Kernel Density Estimation of Toyota Highway Mpg**  den\_T2 <- density(TOYOTA$High\_mpg, bw = 4, na.rm = TRUE)  plot1=plot(den\_T2, main="Kernel Density of TOYOTA Highway\_mpg", ylab = "Density")   1. **R-CODES FOR SAMPLE SIZE DETERMINATION AND ESTIMATIONS**   **library(ggplot2)**  **library(funModeling)**  **library(tidyverse)**  **library(Hmisc)**  **library(dplyr)**  **library(plotly)**  **library(nnet)**  **library(ordinal)**  **library(leaps)**  **library(MASS)**  **library(lubridate)**  **library(survey)**  **library(SDaA)**  **library(magrittr)**  **####################### HONDA VEHICLE MAKE CITY MPG ###################**  **library(stratification)**  **library(PracTools)**  **#Sample Size Estimation#**  **var(HONDA$City\_mpg)**  **library(samplingbook)**  sample.size.mean(e= 0.10 , S=sqrt(57.56079), N= 49662, level= 0.95)  **#Neyman Allocation,Stratified Sampling#**  Stratum=c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J","K")  Nh = c(4961, 7022, 80, 3070, 7106, 1528, 10988,11566,560,1385,1396)  Sh = c(8.422, 8.9867, 3.752, 7.035, 6.732, 4.049, 7.474, 7.173, 10.443, 6.734, 7.807)  sample=strAlloc(n.tot = 15300, Nh = Nh, Sh = Sh, alloc = "neyman")  sample  **#Survey Design Stratified#**  **HONDA$P <- NA**  **#input Stratum weights#**  HONDA$P[HONDA$Class == "Compact Car"] <- 4961/1719  HONDA$P[HONDA$Class == "Large Cars"] <- 7022/2596  HONDA$P[HONDA$Class == "Large Sport Ucars"] <- 80/12  HONDA$P[HONDA$Class == "Midsize Station Wagons"] <- 3070/888  HONDA$P[HONDA$Class == "Small Pickup Truck 2WD"] <- 7106/1968  HONDA$P[HONDA$Class == "Small sport Utility Cars"] <- 1528/254  HONDA$P[HONDA$Class == "Small Station Wagon"] <- 10988/3378  HONDA$P[HONDA$Class == "Standard Pickup Truck 4WD"] <- 11566/3412  HONDA$P[HONDA$Class == "Standard Ucars 4WD"] <- 560/241  HONDA$P[HONDA$Class == "Two Seaters"] <- 1385/384  HONDA$P[HONDA$Class == "Van Cars"] <- 1396/448  head(HONDA)  **########STRATIFIED RANDOM SAMPLING DESIGN FOR HONDA CITY MPG###########**  stat\_stra=svydesign(id=~1, strata=~Class, weights=~P, data=HONDA)  **#Mean Estimate of Str#**  ybar\_str=svymean(~HONDA$City\_mpg, design = stat\_stra)  ybar\_str  **#Variance of the estimated mean**  var\_ybar\_str=svyvar(svymean(~HONDA$City\_mpg, design = stat\_stra))  var\_ybar\_str  **#Confidence Interval of the Estimate mean#**  confint(ybar\_str)  **#Estimated total of Str**  t\_cup\_str=svytotal(~HONDA$City\_mpg, design = stat\_stra)  t\_cup\_str  **#Variance of the estimated total**  var\_t\_cup\_str=svyvar(svytotal(~HONDA$City\_mpg, design = stat\_stra))  var\_t\_cup\_str  **#Design Effect#**  **svymean(~HONDA$City\_mpg, design = stat\_stra, deff = TRUE)**  **#######SIMPLE RANDOM SAMPLING FOR HONDA CITY MPG##**  **####estimate the statistics quantities for SRS ########**  N=49662  n=15300  HONDA$fpc\_srs=N  HONDA$wgt\_stra=N/n  srs\_design=svydesign(ids=~1, weights =~wgt\_stra, data=HONDA)  **#Estimates SRS#**  **#Mean Estimate of SRS#**  ybar\_srs=svymean(~HONDA$City\_mpg,srs\_design)  ybar\_srs  **#Confidence interval of the mean eastimate**  confint(svymean(~HONDA$City\_mpg,srs\_design),df=degf(srs\_design))  **#Total Estimate of SRS#**  t\_cup\_srs=svytotal(~City\_mpg,srs\_design)  t\_cup\_srs  **#Variance of the eastimated total for SRS for Honda City mpg**  var\_t\_cup\_srs=5487.9^2  var\_t\_cup\_srs  **####################### HONDA VEHICLE MAKE HIGHWAY MPG ###################**  **library(stratification)**  **library(PracTools)**  **#Sample Size Estimation#**  var(HONDA$High\_mpg)  library(samplingbook)  sample.size.mean(e= 0.10 , S=sqrt(60.35022), N= 49662, level= 0.95)  **#Neyman Allocation,Stratified Sampling#**  Stratum=c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J","K")  Nh = c(4961, 7022, 80, 3070, 7106, 1528, 10988,11566,560,1385,1396)  Sh = c(8.2381470, 8.6201133, 5.2607786, 7.1932169, 7.1026195, 5.6082737, 7.7528343, 7.6176532, 9.5646146, 7.3558717, 7.7146237)  sample=strAlloc(n.tot = 15806, Nh = Nh, Sh = Sh, alloc = "neyman")  sample  **#Survey Design Stratified#**  HONDA$P <- NA  **#input Stratum weights#**  HONDA$P[HONDA$Class == "Compact Car"] <- 4961/1689  HONDA$P[HONDA$Class == "Large Cars"] <- 7022/2501  HONDA$P[HONDA$Class == "Large Sport Ucars"] <- 80/17  HONDA$P[HONDA$Class == "Midsize Station Wagons"] <- 3070/913  HONDA$P[HONDA$Class == "Small Pickup Truck 2WD"] <- 7106/2085  HONDA$P[HONDA$Class == "Small sport Utility Cars"] <- 1528/354  HONDA$P[HONDA$Class == "Small Station Wagon"] <- 10988/3520  HONDA$P[HONDA$Class == "Standard Pickup Truck 4WD"] <- 11566/3640  HONDA$P[HONDA$Class == "Standard Ucars 4WD"] <- 560/221  HONDA$P[HONDA$Class == "Two Seaters"] <- 1385/421  HONDA$P[HONDA$Class == "Van Cars"] <- 1396/445  head(HONDA)  **########STRATIFIED RANDOM SAMPLING DESIGN FOR HONDA HIGHWAY MPG###########**  stat\_stra=svydesign(id=~1, strata=~Class, weights=~P, data=HONDA)  **#Mean Estimate of Str#**  ybar\_str=svymean(~HONDA$High\_mpg, design = stat\_stra)  ybar\_str  **#Variance of the estimated mean**  var\_ybar\_str=0.035^2  var\_ybar\_str  **#Confidence Interval of the Estimate mean#**  confint(ybar\_str)  **#Estimated total of Stratified sampling**  t\_cup\_str=svytotal(~HONDA$High\_mpg, design = stat\_stra)  t\_cup\_str  **#Variance of the estimated total**  var\_t\_cup\_str=6083.6^2  var\_t\_cup\_str  **#Design Effect#**  **svymean(~HONDA$High\_mpg, design = stat\_stra, deff = TRUE)**  **##############SIMPLE RANDOM SAMPLE FOR HONDA HIGHWAY MPG########**  N=49662  n=15806  HONDA$fpc\_srs=N  HONDA$wgt\_stra=N/n  srs\_design=svydesign(ids=~1, weights =~wgt\_stra, data=HONDA**)**  **#Estimates SRS#**  **#Mean Estimate of SRS#**  ybar\_srs=svymean(~HONDA$High\_mpg,srs\_design)  ybar\_srs  **#Confidence interval of the mean eastimate**  confint(svymean(~HONDA$High\_mpg,srs\_design),df=degf(srs\_design))  **#Total Estimate of SRS#**  t\_cup\_srs=svytotal(~High\_mpg,srs\_design)  t\_cup\_srs  **#Variance of the eastimate**  var\_t\_cup\_srs=5439.4^2  var\_t\_cup\_srs  **####################### TOYOTA VEHICLE MAKE CITY MPG ###################**  **library(stratification)**  **library(PracTools)**  **#Sample Size Estimation#**  var(TOYOTA$City\_mpg)  library(samplingbook)  sample.size.mean(e= 0.10 , S=sqrt(33.29825), N= 49127, level= 0.95)  **#Neyman Allocation,Stratified Sampling#**  Stratum=c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J","K")  Nh = c(7981, 2139, 6578, 908, 3469, 6769, 5233, 7106,3109, 5054,781)  Sh = c(5.1748583, 3.7300218, 6.2753241, 4.4269300, 4.2018262, 6.5574165, 6.5111980, 5.6265052, 5.4679364, 6.0631583, 5.4509082)  sample=strAlloc(n.tot = 10149, Nh = Nh, Sh = Sh, alloc = "neyman")  sample  **#Survey Design Stratified#**  **TOYOTA$P <- NA**  **#input Stratum weights#**  TOYOTA$P[TOYOTA$Class == "Compact Car"] <- 7981/1500  TOYOTA$P[TOYOTA$Class == "Large Sport Utility cars"] <- 2139/290  TOYOTA$P[TOYOTA$Class == "Midsize Car"] <- 6578/1499  TOYOTA$P[TOYOTA$Class == "Midsize Station Wagons"] <- 908/146  TOYOTA$P[TOYOTA$Class == "Small Pickup Truck 2WD"] <- 3469/529  TOYOTA$P[TOYOTA$Class == "Small Standard Wagons"] <- 6769/1612  TOYOTA$P[TOYOTA$Class == "Small Sport Utility Cars"] <- 5233/1237  TOYOTA$P[TOYOTA$Class == "Standard Pickup Truck 4WD"] <- 7106/1452  TOYOTA$P[TOYOTA$Class == "Standard Utility cars 4WD"] <- 3109/617  TOYOTA$P[TOYOTA$Class == "Two Seaters"] <- 5054/1112  TOYOTA$P[TOYOTA$Class == "Van Cars"] <- 781/154  head(TOTOTA)  **########STRATIFIED RANDOM SAMPLING DESIGN FOR TOYOTA CITY MPG###########**  stat\_stra=svydesign(id=~1, strata=~Class, weights=~P, data=TOYOTA)  **#Mean Estimate of Str#**  ybar\_str=svymean(~TOYOTA$City\_mpg, design = stat\_stra)  ybar\_str  **##Variance of the estimated mean**  var\_ybar\_str=0.0251^2  var\_ybar\_str  **#Confidence interval of the mean eastimate**  confint(svymean(~TOYOTA$City\_mpg,stat\_stra),df=degf(stat\_stra))  **#Estimated total of Str**  t\_cup\_str=svytotal(~TOYOTA$City\_mpg, design = stat\_stra)  t\_cup\_str  **#Variance of the estimated total**  var\_t\_cup\_str=6104.5^2  var\_t\_cup\_str  **#Design Effect#**  svymean(~TOYOTA$City\_mpg, design = stat\_stra, deff = TRUE)  **##########SIMPLE RANDOM SAMPLING FOR TOYOTA CITY MPG ########################**  **####estimate the statistics quantities for SRS ########**  N=49127  n=10149  TOYOTA$fpc\_srs=N  TOYOTA$wgt\_stra=N/n  srs\_design=svydesign(ids=~1, weights =~wgt\_stra, data=TOYOTA)  **#Estimates SRS#**  **#Mean Estimate of SRS#**  ybar\_srs=svymean(~TOYOTA$City\_mpg,srs\_design)  ybar\_srs  **#Confidence interval of the mean eastimate**  confint(svymean(~TOYOTA$City\_mpg,srs\_design),df=degf(srs\_design))  **####Variance of the estimated mean**  var\_ybar\_srs=0.026^2  var\_ybar\_srs    **#Total Estimate of SRS#**  t\_cup\_srs=svytotal(~City\_mpg,srs\_design)  t\_cup\_srs  **#Variance of the eastimated total for SRS for TOYOTA City mpg**  var\_t\_cup\_srs=6191.1^2  var\_t\_cup\_srs  **####################### TOYOTA VEHICLE MAKE HIGHWAY MPG ###################**  **library(stratification)**  **library(PracTools)**  **#Sample Size Estimation#**  var(TOYOTA$High\_mpg)  library(samplingbook)  sample.size.mean(e= 0.10 , S=sqrt(42.03938), N= 49127, level= 0.95)  **#Neyman Allocation,Stratified Sampling#**  Stratum=c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J","K")  Nh = c(7981, 2139, 6578, 908, 3469, 6769, 5233, 7106,3109, 5054,781)  Sh = c(6.2014899, 5.2630210, 6.6309589, 5.7864360, 5.5533916, 7.1158215, 7.1021862, 6.3547713, 6.3498139, 6.4573177, 6.1549335)  sample=strAlloc(n.tot = 12154, Nh = Nh, Sh = Sh, alloc = "neyman")  sample  **#Survey Design Stratified#**  TOYOTA$P <- NA  **#input Stratum weights#**  TOYOTA$P[TOYOTA$Class == "Compact Car"] <- 7981/1900  TOYOTA$P[TOYOTA$Class == "Large Sport Utility cars"] <- 2139/432  TOYOTA$P[TOYOTA$Class == "Midsize Car"] <- 6578/1675  TOYOTA$P[TOYOTA$Class == "Midsize Station Wagons"] <- 908/202  TOYOTA$P[TOYOTA$Class == "Small Pickup Truck 2WD"] <- 3469/740  TOYOTA$P[TOYOTA$Class == "Small Standard Wagons"] <- 6769/1849  TOYOTA$P[TOYOTA$Class == "Small Sport Utility Cars"] <- 5233/1427  TOYOTA$P[TOYOTA$Class == "Standard Pickup Truck 4WD"] <- 7106/1734  TOYOTA$P[TOYOTA$Class == "Standard Utility cars 4WD"] <- 3109/758  TOYOTA$P[TOYOTA$Class == "Two Seaters"] <- 5054/1253  TOYOTA$P[TOYOTA$Class == "Van Cars"] <- 781/184  head(TOTOTA)  **########STRATIFIED RANDOM SAMPLING DESIGN FOR TOYOTA HIGHWAY MPG###########**  stat\_stra=svydesign(id=~1, strata=~Class, weights=~P, data=TOYOTA)  **#Mean Estimate of Str#**  ybar\_str=svymean(~TOYOTA$High\_mpg, design = stat\_stra)  ybar\_str  **#Variance of the estimated mean**  var\_ybar\_str=0.0288^2  var\_ybar\_str  **#Confidence Interval of the Estimate mean#**  confint(ybar\_str)  **#Estimated total of Stratified sampling**  t\_cup\_str=svytotal(~TOYOTA$High\_mpg, design = stat\_stra)  t\_cup\_str  **#Variance of the estimated total**  var\_t\_cup\_str=7014^2  var\_t\_cup\_str  **#Design Effect#**  svymean(~TOYOTA$High\_mpg, design = stat\_stra, deff = TRUE)  **##############SIMPLE RANDOM SAMPLE FOR TOYOTA HIGHWAY MPG########**  N=49127  n=10149  TOYOTA$fpc\_srs=N  TOYOTA$wgt\_stra=N/n  srs\_design=svydesign(ids=~1, weights =~wgt\_stra, data=TOYOTA)  **#Estimates SRS#**  **#Mean Estimate of SRS#**  ybar\_srs=svymean(~TOYOTA$High\_mpg,srs\_design)  ybar\_srs  var\_ybar\_srs=0.0293^2  var\_ybar\_srs    **#Confidence interval of the mean eastimate**  confint(svymean(~TOYOTA$High\_mpg,srs\_design),df=degf(srs\_design))  **#Total Estimate of SRS#**  t\_cup\_srs=svytotal(~High\_mpg,srs\_design)  t\_cup\_srs  **#Variance of the eastimate**  var\_t\_cup\_srs=6956.4^2  var\_t\_cup\_srs  **####Gain from Stratification#####**  **# The Relative Gain From Stratification For Honda City Mpg**  Gain\_from\_str=var\_t\_cup\_str/var\_t\_cup\_srs  Gain\_from\_str=3038523/30117046  Gain\_from\_str  **# The Relative Gain From Stratification For Honda Highway Mpg**  Gain\_from\_str=var\_t\_cup\_str/var\_t\_cup\_srs  Gain\_from\_str=3021234/29587072  Gain\_from\_str  **# The Relative Gain From Stratification For TOYOTA City Mpg**  Gain\_from\_str=var\_t\_cup\_str/var\_t\_cup\_srs  Gain\_from\_str=37264920/38329719  Gain\_from\_str  **# The Relative Gain From Stratification For TOYOTA City Mpg**  Gain\_from\_str=var\_t\_cup\_str/var\_t\_cup\_srs  Gain\_from\_str=2001822/48391501  Gain\_from\_str | |  | |  | | |  | |