libname IPUMS "C:\Users\Emeka\Desktop\ECON 453\PROJECT";

filename ten "usa\_00009.dat";

proc format cntlout = IPUMS.usa\_00009\_f;

value YEAR\_f

1850 = "1850"

1860 = "1860"

1870 = "1870"

1880 = "1880"

1900 = "1900"

1910 = "1910"

1920 = "1920"

1930 = "1930"

1940 = "1940"

1950 = "1950"

1960 = "1960"

1970 = "1970"

1980 = "1980"

1990 = "1990"

2000 = "2000"

2001 = "2001"

2002 = "2002"

2003 = "2003"

2004 = "2004"

2005 = "2005"

2006 = "2006"

2007 = "2007"

2008 = "2008"

2009 = "2009"

2010 = "2010"

;

value GQ\_f

0 = "Vacant unit"

1 = "Households under 1970 definition"

2 = "Additional households under 1990 definition"

3 = "Group quarters--Institutions"

4 = "Other group quarters"

5 = "Additional households under 2000 definition"

6 = "Fragment"

;

value NCHILD\_f

0 = "No Children"

1 = "1 child"

2,3 = "2 to 3 Children"

4-9 = "4 or more Children"

;

value AGE\_f

000 = "Less than 1 year old"

001 = "1"

002 = "2"

003 = "3"

004 = "4"

005 = "5"

006 = "6"

007 = "7"

008 = "8"

009 = "9"

010 = "10"

011 = "11"

012 = "12"

013 = "13"

014 = "14"

015 = "15"

016 = "16"

017 = "17"

048-066="Baby Boomers"

030-047="Generation X"

018-029="Generation Y"

066 = "66"

067 = "67"

068 = "68"

069 = "69"

070 = "70"

071 = "71"

072 = "72"

073 = "73"

074 = "74"

075 = "75"

076 = "76"

077 = "77"

078 = "78"

079 = "79"

080 = "80"

081 = "81"

082 = "82"

083 = "83"

084 = "84"

085 = "85"

086 = "86"

087 = "87"

088 = "88"

089 = "89"

090 = "90 (90+ in 1980 and 1990)"

091 = "91"

092 = "92"

093 = "93"

094 = "94"

095 = "95"

096 = "96"

097 = "97"

098 = "98"

099 = "99"

100 = "100 (100+ in 1970)"

101 = "101"

102 = "102"

103 = "103"

104 = "104"

105 = "105"

106 = "106"

107 = "107"

108 = "108"

109 = "109"

110 = "110"

111 = "111"

112 = "112 (112+ in the 1980 internal data)"

113 = "113"

114 = "114"

115 = "115 (115+ in the 1990 internal data)"

116 = "116"

117 = "117"

118 = "118"

119 = "119"

120 = "120"

121 = "121"

122 = "122"

123 = "123"

124 = "124"

125 = "125"

126 = "126"

129 = "129"

130 = "130"

135 = "135"

;

value SEX\_f

1 = "Male"

2 = "Female"

;

value MARST\_f

1,2 = "Married"

3 = "Separated"

4 = "Divorced"

5 = "Widowed"

6 = "Single"

;

value RACESING\_f

1 = "White"

2 = "Black"

3 = "Native American"

4 = "Asian/Pacific Islander"

5 = "Other race"

;

value RACESINGD\_f

10 = "White"

12 = """Other race"", Hispanic"

20 = "Black"

21 = "Mulatto"

30 = "AI (American Indian)"

31 = "AN (Alaskan Native)"

32 = "AI/AN (American Indian/Alaskan Native)"

40 = "Asian Indian"

41 = "Chinese"

42 = "Filipino"

43 = "Japanese"

44 = "Korean"

45 = "Asian "

46 = "Hawaiian"

47 = "PI (Pacific Islander)"

48 = "Asian and PI (Pacific Islander)"

50 = "Other race, non-Hispanic"

;

value HCOVANY\_f

1 = "No health insurance coverage"

2 = "With health insurance coverage"

;

value HCOVPRIV\_f

1 = "0"

2 = "1"

;

value HINSEMP\_f

1 = "No insurance through employer/union"

2 = "Has insurance through employer/union"

;

value EDUC\_f

00,01,02,03,04,05,06 = "HS Diploma or Less"

07 = "1 year of college"

08,09 = "2 years of college"

10 = "4 years of college"

11 = "5+ years of college"

;

value EDUCD\_f

000 = "N/A or no schooling"

001 = "N/A"

002 = "No schooling completed"

010 = "Nursery school to grade 4"

011 = "Nursery school, preschool"

012 = "Kindergarten"

013 = "Grade 1, 2, 3, or 4"

014 = "Grade 1"

015 = "Grade 2"

016 = "Grade 3"

017 = "Grade 4"

020 = "Grade 5, 6, 7, or 8"

021 = "Grade 5 or 6"

022 = "Grade 5"

023 = "Grade 6"

024 = "Grade 7 or 8"

025 = "Grade 7"

026 = "Grade 8"

030 = "Grade 9"

040 = "Grade 10"

050 = "Grade 11"

060 = "Grade 12"

061 = "12th grade, no diploma"

062 = "High school graduate or GED"

063 = "Regular high school diploma"

064 = "GED or alternative credential"

065 = "Some college, but less than 1 year"

070 = "1 year of college"

071 = "1 or more years of college credit, no degree"

080 = "2 years of college"

081 = "Associate's degree, type not specified"

082 = "Associate's degree, occupational program"

083 = "Associate's degree, academic program"

090 = "3 years of college"

100 = "4 years of college"

101 = "Bachelor's degree"

110 = "5+ years of college"

111 = "6 years of college (6+ in 1960-1970)"

112 = "7 years of college"

113 = "8+ years of college"

114 = "Master's degree"

115 = "Professional degree beyond a bachelor's degree"

116 = "Doctoral degree"

;

value CLASSWKR\_f

0 = "N/A"

1 = "Self-employed"

2 = "Works for wages"

;

value CLASSWKRD\_f

00 = "Not Working"

10 = "Self-employed"

11 = "Employer"

12 = "Working on own account"

13,14,29 = "Self-Employed"

20 = "Works for wages"

21 = "Works on salary (1920)"

22,23 = "Wage/salary, private & Non-Profit"

24 = "Wage/salary, government"

25,27,28 = "Government Employee"

26 = "Armed forces"

;

value UHRSWORK\_f

00 = 'None'

1 - 35 = 'Part-Time'

36 - 40 = 'Full-Time'

41 - 99='More than 40';

;

run;

data IPUMS.ten;

infile "C:\Users\Emeka\Desktop\ECON 453\PROJECT\usa\_00009.dat" pad missover lrecl=61;

input

YEAR 1-4

DATANUM 5-6

SERIAL 7-14

HHWT 15-24 .2

GQ 25-25

PERNUM 26-29

PERWT 30-39 .2

NCHILD 40-40

AGE 41-43

SEX 44-44

MARST 45-45

RACESING 46-46

RACESINGD 47-48

HCOVANY 49-49

HCOVPRIV 50-50

HINSEMP 51-51

EDUC 52-53

EDUCD 54-56

CLASSWKR 57-57

CLASSWKRD 58-59

UHRSWORK 60-61

;

label

YEAR = "Census year"

DATANUM = "Data set number"

SERIAL = "Household serial number"

HHWT = "Household weight"

GQ = "Group quarters status"

PERNUM = "Person number in sample unit"

PERWT = "Person weight"

NCHILD = "Number of children in household"

AGE = "Age"

SEX = "Sex"

MARST = "Marital status"

RACESING = "Racial identification"

RACESINGD = "Race: Single race identification [detailed version]"

HCOVANY = "Any health insurance coverage"

HCOVPRIV = "Private health Insurance(PHI)"

HINSEMP = "Health insurance through employer/union"

EDUC = "Educational attainment"

EDUCD = "Educational attainment [detailed version]"

CLASSWKR = "Class of worker [general version]"

CLASSWKRD = "Class of worker [detailed version]"

UHRSWORK = "Usual hours worked per week"

;

format

YEAR YEAR\_f.

GQ GQ\_f.

NCHILD NCHILD\_f.

AGE AGE\_f.

SEX SEX\_f.

MARST MARST\_f.

RACESING RACESING\_f.

RACESINGD RACESINGD\_f.

HCOVANY HCOVANY\_f.

HCOVPRIV HCOVPRIV\_f.

HINSEMP HINSEMP\_f.

EDUC EDUC\_f.

EDUCD EDUCD\_f.

CLASSWKR CLASSWKR\_f.

CLASSWKRD CLASSWKRD\_f.

UHRSWORK UHRSWORK\_f.

;

format

HHWT 11.2

PERWT 11.2

;

**FROM HERE**

**PROJECT ONE**

data one;

set ipums.ten;

if 18 <= age <= 64 then output;

else delete;

run;

ods pdf file= 'C:\Users\Emeka\Desktop\project.pdf';

\*Dependent Varible: Private health insurance coverage;

proc freq data=one;

table hcovpriv;

title 'Private health insurance coverage';

run;

proc gchart data=one;

pie hcovpriv/ type=percent value=inside;

title 'Private health insurance coverage';

run; quit;

\*-------------------------------------;

\*Independent Variable: Age;

proc means data=one min max mean;

var age;

title 'Age Distribution';

run;

proc gchart data=one;

vbar age/ discrete type=percent;

title 'Age Distribution';

run;

\*According to www.socialsecurity.gov, the Returement age for people born in 1960 and over is 67. People born before 1960 are eligibe by age 65.;

\*-------------------------------------;

\*Independent Variable: Marital Status;

proc freq data=one;

table MARST;

title 'Marital Status';

run;

Proc gchart data=one;

vbar MARST/ discrete type=percent;

title 'Marital Status';

run; quit;

\*-------------------------------------;

\*Independent Variable: Education;

proc freq data=one;

table educ;

title 'Education Attainment';

run;

proc gchart data=one;

vbar educ/ discrete type=percent;

title 'Education Attainment';

run; quit;

\*-------------------------------------;

\*Independent Variable: Sex;

proc freq data=one;

table sex;

title 'Sex';

run;

proc gchart data=one;

pie sex/ discrete type=percent value=inside; \*how do i remove frequency from around the chart?;

title 'Distribution of Sex';

run; quit;

\*-------------------------------------;

\*Independent Variable: Classification of Work;

proc freq data=one;

table classwkrd;

title 'Classification of Work';

run;

proc gchart data=one;

vbar classwkrd/ discrete type=percent;

title 'Classification of Work';

run; quit;

\*-------------------------------------;

\*Independent Variable: Work hours Per week;

proc freq data=one;

table uhrswork;

title 'Hours Worked Per Week';

run;

proc gchart data=one;

vbar uhrswork/ discrete nostats type=percent;

title 'Hours Worked Per Week';

run; quit;

\*-------------------------------------;

\*Independent Variable: Race;

proc freq data=one;

table racesing;

title 'Race';

run;

proc gchart data=one;

vbar racesing/ discrete type=percent;

title 'Race';

run;

\*-------------------------------------;

\*Independent Variable: Number of Children;

proc freq data=one;

table nchild;

title 'Number of Children in Household';

run;

proc gchart data=one;

pie nchild/ discrete type=percent value=inside;

title 'Number of Children in Household';

run;

\*---------------------------------------------------------------------------------------------------------------;

\*Dependent Varible: Private health insurance coverage;

proc freq data=ipums.ten;

table hcovpriv;

title 'Private health insurance coverage';

run;

proc gchart data=one;

pie hcovpriv/ type=percent value=inside;

title 'Private health insurance coverage';

run; quit;

\*-----------------------------------------------------------------------------------------------------------------;

\*-----------------------------------------------------------------------------------------------------------------;

\*-----------------------------------------------------------------------------------------------------------------;

\*-----------------------------------------------------------------------------------------------------------------;

\*-----------------------------------------------------------------------------------------------------------------;

\*-----------------------------------------------------------------------------------------------------------------;

data two;

set one;

if hcovpriv = 2 then care =1;

if hcovpriv = 1 then care =0;

run;

\*COnditional Statistics: AGE;

proc gchart data=two;

vbar age/sumvar = Care type=mean discrete missing raxis=axis1;

format Care percent8.;

run;

proc tabulate data=two;

var Care;

class age;

table age all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Marital Status;

proc gchart data=two;

vbar marst/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class marst;

table marst all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Education Attainment;

proc gchart data=two;

vbar educ/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class educ;

table educ all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Sex;

proc gchart data=two;

vbar sex/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class sex;

table sex all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Weekly Work Hours;

proc gchart data=two;

vbar uhrswork/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class uhrswork;

table uhrswork all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Race;

proc gchart data=two;

vbar racesing/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class racesing;

table racesing all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

\*COnditional Statistics: Children in Hourshold;

proc gchart data=two;

vbar nchild/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class nchild;

table nchild all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

ods pdf file= 'C:\Users\Emeka\Desktop\class.pdf';

proc gchart data=two;

vbar classwkrd/sumvar = Care type=mean discrete missing;

format Care percent6.1;

run;

proc tabulate data=two;

var Care;

class classwkrd;

table classwkrd all='Overall', Care\*mean='Percent'\*f=percent8.1;

run;

ods pdf close;

\*----------------------------------------------;

data two;

set one;

if hcovpriv = 2 then care =1;

if hcovpriv = 1 then care =0;

if 030<= age >=047 then generationx = 1;

else generationx = 0;

if 018<= age >=029 then generationy = 1;

else generationy = 0;

if marst = 3 then marriagesep = 1;

else marriagesep = 0;

if marst = 4 then marriagediv = 1;

else marriagediv = 0;

if marst = 5 then marriagewid = 1;

else marriagewid = 0;

if marst = 6 then marriagesing = 1;

else marriagesing = 0;

if educ = 07 then school1 = 1;

else school1 = 0;

if educ = 08 & 09 then school2 = 1;

else school2 = 0;

if educ = 10 then school4 = 1;

else school4 = 0;

if educ = 11 then school5 = 1;

else school5 = 0;

if Sex = 1 then genderm = 1;

else genderm = 0;

if classwkrd = 00 then workclassNot = 1;

else workclassNot = 0;

if classwkrd = 13 then workclassself = 1;

else workclassself = 0;

if classwkrd = 14 then workclassself = 1;

else workclassself = 0;

if classwkrd = 29 then workclassself = 1;

else workclassself = 0;

if classwkrd = 25 then workclassgovt = 1;

else workclassgovt = 0;

if classwkrd = 27 then workclassgovt = 1;

else workclassgovt = 0;

if classwkrd = 28 then workclassgovt = 1;

else workclassgovt = 0;

if uhrswork = 00 then hoursnone = 1;

else hoursnone = 0;

if 1<= uhrswork >=35 then hourspart = 1;

else hourspart = 0;

if 41<= uhrswork >=99 then hoursover = 1;

else hoursover = 0;

if racesing = 2 then raceblk = 1;

else raceblk = 0;

if racesing = (3 & 5) then racenatothr = 1;

else racenatothr = 0;

if racesing = 4 then raceAsn = 1;

else raceAsn = 0;

if nchild = 0 then Children1 = 1;

else Children1 = 0;

if nchild = 2 then Children2or3 = 1;

else Children2or3 = 0;

if nchild = 3 then Children2or3 = 1;

else Children2or3 = 0;

if nchild = 4 then Children4to9 = 1;

else Children4to9 = 0;

if nchild = 5 then Children4to9 = 1;

else Children4to9 = 0;

if nchild = 6 then Children4to9 = 1;

else Children4to9 = 0;

if nchild = 7 then Children4to9 = 1;

else Children4to9 = 0;

if nchild = 8 then Children4to9 = 1;

else Children4to9 = 0;

if nchild = 9 then Children4to9 = 1;

else Children4to9 = 0;

run;

proc logistic data=two descending;

model care =

Children4to9 Children2or3 Children1

raceAsn racenatothr raceblk

hoursover hourspart hoursnone

workclassgovt workclassself workclassNot

genderm school5 school4 school2 school1

marriagesing marriagewid marriagediv

marriagesep generationx generationy;

output out= three pred=prob;

run;

proc univariate data=three;

var prob;

histogram prob/kernel nobars (lower=0 upper=1 fill);

run;

proc means data=three n mean min median max;

var prob;

run;

proc sort data=three;

by care;

run;

proc print data=three (obs=10);

var prob Children4to9 Children2or3 Children1

raceAsn racenatothr raceblk

hoursover hourspart hoursnone

workclassgovt workclassself workclassNot

genderm school5 school4 school2 school1

marriagesing marriagewid marriagediv

marriagesep generationx generationy;

where prob<0.90;

run;

proc print data=three (obs=20);

var care prob

Children4to9 Children2or3 Children1

raceAsn racenatothr raceblk

hoursover hourspart hoursnone

workclassgovt workclassself workclassNot

genderm school5 school4 school2 school1

marriagesing marriagewid marriagediv

marriagesep generationx generationy;

format prob percent30.10;

where prob < 0.000001;

run;

proc print data=three (obs=20);

var care prob;

format prob percent12.4;

where 0.35 < prob <0.65;

run;

proc print data=three (obs=20);

var care prob;

format prob percent12.4;

where 0.90 < prob <1.1;

run;

ods pdf close;

proc univariate data=three;

var prob;

histogram prob/KERNEL nobars fill;

run;

proc univariate data=negatives;

var estimate;

histogram estimate/kernel nobars;

run;

proc univariate data=positives;

var estimate;

histogram estimate/kernel nobars;

run;

ods pdf file= 'C:\Users\Emeka\Desktop\regression.pdf';;

proc means data=negatives n min median max;

var estimate;

run;

proc means data=positives n min median max;

var estimate;

run;

ods pdf close;

proc contents data=three;

run;

**PROJECT 2**

PROC IMPORT OUT= WORK.white   
            DATAFILE= "\\Client\C$\Users\Emeka Ume\Desktop\winequality-white.csv"   
            DBMS=CSV REPLACE;  
     GETNAMES=YES;  
     DATAROW=2;   
  
  
data white1;  
set white;  
if quality>=6 then qual=1; else qual=0;  
run;  
ods html close;  
ods html;  
run;  
  
  
proc logistic data=white1;  
model qual=fixed\_acidity volatile\_acidity residual\_sugar ph sulphates density chlorides free\_sulfur\_dioxide  
citric\_acid total\_sulfur\_dioxide alcohol / lackfit ;  
run;  
  
/\*Remove Citric Acid\*/  
proc logistic data=white1;  
model qual=fixed\_acidity volatile\_acidity residual\_sugar ph sulphates density chlorides free\_sulfur\_dioxide  
total\_sulfur\_dioxide alcohol / lackfit ;  
run;  
  
/\*Remove Citric Acid & Fixed\_acidity\*/  
proc logistic data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates density chlorides free\_sulfur\_dioxide  
total\_sulfur\_dioxide alcohol / lackfit ;  
run;  
  
/\*Remove Citric Acid & Fixed\_acidity & chlorides\*/  
proc logistic data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates density  free\_sulfur\_dioxide  
alcohol Total\_sulfur\_dioxide/ lackfit ;  
run;  
  
/\*Remove Citric Acid & Fixed\_acidity & Total\_sulfur\_dioxide & chlorides\*\*\*\*\*\*\*\*/  
proc logistic data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates free\_sulfur\_dioxide alcohol density / lackfit ;  
run;  
  
/\*VIF & Tol test with Density\*/  
proc reg data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates density free\_sulfur\_dioxide alcohol / vif tol ;  
run;  
  
  
/\*VIF & Tol test without Density\*/  
proc reg data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates free\_sulfur\_dioxide alcohol / vif tol;  
run;  
  
/\*Remove Citric Acid & Fixed\_acidity & Total\_sulfur\_dioxide & chlorides Density\*/  
proc logistic data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates free\_sulfur\_dioxide alcohol/ lackfit;  
run;  
  
  
/\*\*\*NEW\*\*\*/  
proc genmod data=white1;  
model qual= volatile\_acidity residual\_sugar ph sulphates free\_sulfur\_dioxide alcohol/ obstats influence residuals d=b;  
run;  
  
  
proc freq data=white;  
table quality;  
run;