New York City Health and Nutrition Examination Survey (NYC HANES) 2004 and 2013-14:

Part III – Data analysis using SAS and SUDAAN

Disclosure

- All material included in this tutorial presentation is available for public use as an educational tool.
- Results from this tutorial analysis **should not** be used as official estimates or references for research.

Acknowledgement and Contact Info

- Created by NYC HANES 2013-14 Team:

  - Claudia Chernov (cchernov@health.nyc.gov)
  - Rania Kanchi (rania.kanchi@nyumc.org)
  - Lorna Thorpe (lorna.thorpe@nyumc.org)
  - Sharon Perlman (sperlma1@health.nyc.gov)

Thanks to Jesica Rodriguez-Lopez for her contribution in creating NYC HANES training material

Purpose

- This tutorial serves as a reference for standard analyses of NYC HANES 2004 and 2013-14 data
- Presented here is a step-by-step analysis of one health indicator (diabetes)
- SAS syntax included in this presentation can be found as sample code. To gain access to the .sas file, enter your email address on the DATA tab of our website: http://nychanes.org/data/

Outline

- This tutorial presentation will cover:
  – Recoding of variables
  – Weight adjustment due to non-response
  – Age standardization
  – Prevalence estimate calculations
  – Stratified analysis (prevalence estimates by gender)
  – Comparison across years

Objective

- Calculate the prevalence of diabetes in NYC residents, overall and by gender, and compare prevalence across survey years
  - Diabetes defined as:
    - Diagnosed diabetes: if participant was ever told by a doctor or health care professional that he or she has diabetes.
    - Undiagnosed diabetes: if participant has either fasting plasma glucose (FPG) of ≥126 mg/dl or A1c of ≥6.5% without having been previously diagnosed with diabetes.
  - Total diabetes = sum of diagnosed & undiagnosed diabetes.

PREVALENCE ESTIMATES

Analytic checklist for prevalence estimates

1. Call in dataset
2. Identify outcome(s) of interest; recode if necessary
3. Identify appropriate weight
4. Create a new weight that has been adjusted for non-response (SAS)
5. Test the new weight
6. Run prevalence estimates (SUDAAN or SAS Proc Survey)

1. Call in dataset NYC HANES 2013-14

```
/*Download datasets, and save on your own computer*/
/*For this tutorial, we will assume that dataset has been renamed
    NYCHANES_2013-14 and has been saved in C:\Data\2014*/

libname NYCH14 'C:\Data\2014';

data NYCHANES14;
        set NYCH14.NYCHANES_2013-14;
run;
```

Health

NYC hanes Health & Nutrition Examination Survey

NYC HANES New York City Health and Nutrition Examination Survey

CUNY SPH

Step 2: Identify outcome of interest

1. Identify outcome of interest
   - Diabetes: Diabetes based on FPG, A1c, or self-reported diagnosis:
     - Fasting Plasma Glucose (FPG) equal to or greater than 126 mg/dL, or
     - A1c equal to or greater than 6.5% (regardless of fasting status), or
     - Self-reported diagnosis of diabetes

Step 2: Identify outcome of interest (continued)

2. Variables needed for recode

Variable	2004	2013-14
Fasting Indicator	WTSF1F	DBT_Fasting
Glucose	LBXGLU	Glucose
A1c	LBXGH	A1c
Self-Reported Diabetes	DIQ010	DIQ_1

Health

NYC hanes Health & Nutrition Examination Survey

NYC HANES New York City Health and Nutrition Examination Survey

CUNY SPH

Step 2 continued: Recode/create outcome of interest NYC HANES 2013-14

```
data dbts14;
set NYCHANES14;

*Define diabetes using FPG or A1c or previous diagnosis of diabetes;

/*Diabetic: Fasted and have plasma glucose =>126 or A1C =>6.5 or self-reported diagnosis*/
if DBT_Fasting=1 and (Glucose >=126 or A1c >=6.5) or DIQ_1=1 then dbts=1;

/*Diabetic: Did not fast and have A1c =>6.5 or self-reported diagnosis*/
else if DBT_Fasting=0 and (A1c =>6.5 or DIQ_1=1) then dbts=1;

/*Non-Diabetic: Fasted and have plasma glucose<126 and A1c<6.5 and no self-reported diagnosis*/
else if DBT_Fasting=1 and (0<Glucose<126 or 0<A1c<6.5)  and DIQ_1 ne 1 then dbts=2;

/*Non-Diabetic: Did not fast and have A1c<6.5 and no self-reported diagnosis*/
else if DBT_Fasting=0 and 0<A1c<6.5  and DIQ_1 ne 1 then dbts=2;

label dbts = ‘Diabetes based on FPG, A1c, or Self-Report’;
```

Step 3: Identify appropriate weight (1)

1. Outcome = Diabetes (calculated from blood tests and self-reported diagnosis)
   - Fasting plasma glucose comes from fasting subset
     - Participants who fasted at least 8 hours and have valid glucose laboratory test results
   - A1c comes from blood subset
     - Participants with valid A1c laboratory test results
   - Self-reported diagnosis comes from survey (CAPI)
     - Participants who answered DIQ_1 of the CAPI survey

Step 3: Identify appropriate weight (2)

2. Is our definition of the outcome exclusive or inclusive?
   - If EXCLUSIVE, we define diabetes among ONLY those participants who fasted and who had a valid glucose value.
     - Then use:
       - 2004: Fasting weight (**WTSF1F**)
       - 2013-14: Blood weight (**BLOOD_WT**) and adjust for fasting
     - This definition excludes participants who did not fast as well as participants who did not have a valid glucose measurement, even those with a valid A1c and even those with a diabetes diagnosis.

Step 3: Identify appropriate weight (3)

2. Is our definition of the outcome inclusive or exclusive?
   – If INCLUSIVE, we define diabetes among participants with a valid A1c value OR participants who reported a previous diabetes diagnosis OR participants who both fasted and had a valid plasma glucose value.
     - Then use:
       – 2004: Clinic + Home (**WTSF1CH**)
       – 2013-14: CAPI weight (**CAPI_WT**)

NYC Health · NYC hanes Health & Nutrition Examination Survey · NYC HANES New York City Health and Nutrition Examination Survey · CUNY SPH

Step 3: Identify appropriate weight (4)

- Number of cases to be included in the 2013-14 analysis: **1246**

```
proc freq data=dbts14;
      tables dbts/list;
Run;
```

Diabetes based on FPG, A1c, or Self-Report				
dbts	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1: Yes	192	15.41	192	15.41
2: No	1054	84.59	1246	100
Frequency Missing = 281				

Step 4: Adjust for non-response NYC HANES 2013-14 (1)

1. Set the initial weight variable of the outcome. This variable will be adjusted for non-response.

```
data dbts14;
set dbts14;
 /*1. Initial Diabetes weight. This is the first crucial step in adjusting for non-response. Set the weight to zero for cases that will not be included in the analysis (that is, cases with a missing value for the dbts variable).*/
if dbts=. then dbts_Wt=0;
/*2. Set the weight to the identified appropriate weight for cases that will be included in the analysis. (We will use CAPI_WT, because our definition is inclusive.)*/
else dbts_Wt =CAPI_WT;
run;
```

Step 4: Adjust for non-response NYC HANES 2013-14 (2)

2. Create a new dataset called **samplewts**, in which you sum the initial diabetes weight by age, sex, race/ethnicity.
3. Create a new variable called **samplewt** (this will be the sum of the initial diabetes weight for all observations by age, sex, and race/ethnicity).

```
proc summary data=dbts14 nway;
    var dbts_Wt;
    class Agegrp4c gender race;
    output out=samplewts sum=samplewt;
run;
```

Step 4: Adjust for non-response NYC HANES 2013-14 (3)

Partial printout of samplewts

AGEGRP4C	GENDER	RACE	_TYPE_	_FREQ_	samplewt
1: 20-34	1 =Male	1: Non-Hispanic White	7	85	255196.7
1: 20-34	1 =Male	2: Non-Hispanic Black	7	43	132567.89
1: 20-34	1 =Male	3: Hispanic	7	57	213555.6
1: 20-34	1 =Male	4: Asian	7	37	137913.24
1: 20-34	1 =Male	5: Other	7	11	16415.44
1: 20-34	2: Female	1: Non-Hispanic White	7	95	297919.75
1: 20-34	2: Female	2: Non-Hispanic Black	7	66	166335.22
1: 20-34	2: Female	3: Hispanic	7	81	228598.94
1: 20-34	2: Female	4: Asian	7	59	164521.19
1: 20-34	2: Female	5: Other	7	18	21349.21
2: 35-49	1 =Male	1: Non-Hispanic White	7	61	235447.32
2: 35-49	1 =Male	2: Non-Hispanic Black	7	38	146063.27
2: 35-49	1 =Male	3: Hispanic	7	42	164492.92
2: 35-49	1 =Male	4: Asian	7	25	90891.56
2: 35-49	1 =Male	5: Other	7	10	17850.15
2: 35-49	2: Female	1: Non-Hispanic White	7	60	230650.1
2: 35-49	2: Female	2: Non-Hispanic Black	7	65	172218.62
2: 35-49	2: Female	3: Hispanic	7	58	230223.09
2: 35-49	2: Female	4: Asian	7	29	86412.8
2: 35-49	2: Female	5: Other	7	9	14395.82

Control totals: NYC HANES 2013-14

1. These datasets contain reference totals for each of the five survey weights.
2. The weights in each dataset sum to **6,285,749** and are stratified by age, gender, and race/ethnicity.

Weight	Control dataset name
CAPI	Control_CAPI
Exam	Control_EXAM
Blood	Control_BLOOD
Urine	Control_URINE
Saliva	Control_SALIVA

Step 4: Adjust for non-response NYC HANES 2013-14 (4)

4. Call in the appropriate control total dataset

```
data Control_capi;
    set NYCH14.Control_capi;
run;
```

5. Sort the control total dataset and the samplewts dataset by the variables on which they will merge: Agegrp4c gender race

```
proc sort data=Control_capi;
    by Agegrp4c gender race;
run;
proc sort data=samplewts;
    by Agegrp4c gender race;
run;
```

Step 4: Adjust for non-response NYC HANES 2013-14 (5)

6. Merge the dataset that contains weight control totals with the dataset of summed weights (from the analytic dataset), and create weight factor

```
data wtfactor;
merge samplewts Control_capi;
by Agegrp4c gender race;
wt_factor_dbts=control_totals/samplewt; /*weight factor*/
keep Agegrp4c gender race wt_factor_dbts;
run;
```

Step 4: Adjust for non-response NYC HANES 2013-14 (6)

Partial print out of wtfactor dataset

AGEGRP4C	GENDER	RACE	wt_factor_dbts
1: 20-34	1 =Male	1: Non-Hispanic White	1.31291
1: 20-34	1 =Male	2: Non-Hispanic Black	1.21212
1: 20-34	1 =Male	3: Hispanic	1.25018
1: 20-34	1 =Male	4: Asian	1.27053
1: 20-34	1 =Male	5: Other	1.45349
1: 20-34	2: Female	1: Non-Hispanic White	1.13253
1: 20-34	2: Female	2: Non-Hispanic Black	1.31455
1: 20-34	2: Female	3: Hispanic	1.28022
1: 20-34	2: Female	4: Asian	1.35181
1: 20-34	2: Female	5: Other	1.34763
2: 35-49	1 =Male	1: Non-Hispanic White	1.21079
2: 35-49	1 =Male	2: Non-Hispanic Black	1.31728
2: 35-49	1 =Male	3: Hispanic	1.29863
2: 35-49	1 =Male	4: Asian	1.29986
2: 35-49	1 =Male	5: Other	1.25584
2: 35-49	2: Female	1: Non-Hispanic White	1.07935
2: 35-49	2: Female	2: Non-Hispanic Black	1.43185
2: 35-49	2: Female	3: Hispanic	1.09627
2: 35-49	2: Female	4: Asian	1.2473
2: 35-49	2: Female	5: Other	1.11508

Step 4: Adjust for non-response NYC HANES 2013-14 (7)

7. Sort the new weight factor dataset and your analytic dataset, and then merge the two datasets. Create the weight for your analysis by multiplying the weights of participants with non-missing data by the weight factor.

```
proc sort data=wtfactor; by Agegrp4c gender race; run;
proc sort data=dbts14; by Agegrp4c gender race; run;
data dbts14_2;
merge dbts14 (in=a) wtfactor;
by Agegrp4c gender race;
if a;
/*Adjusted Diabetes Weight*/
dbts_wt_new = dbts_Wt * wt_factor_dbts;
drop dbts_Wt wt_factor_dbts;
/*Adjusted Diabetes Weight Indicator (wt_dbts)*/
if dbts_wt_new = 0 then wt_dbts = 2;
        else wt_dbts = 1;
label   dbts_wt_new = 'Diabetes Weight'
        wt_dbts = 'Diabetes Weight Indicator';
format wt_dbts TE_3F.;
run;
```

Step 5: Test your new weight!
NYC HANES 2013-14

1. Make sure your adjusted weight is defined for all the cases included in the analysis: **1246** (see slide 16)

```
proc freq data = dbts14_2;
table wt_dbts;
run;
```

Diabetes Weight Indicator				
wt_dbts	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1: Yes	1246	81.6	1246	81.6
2: No	281	18.4	1527	100

2. Make sure the total weighted frequency is the total population represented by NYC HANES, **6,285,749**, and the gender distribution matches that of the survey, 46.9% male and 53.1% female:

```
proc freq data = dbts14_2;
table gender;
weight dbts_wt_new;
run;
```

Gender				
Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1=Male	2931913	46.64	2931913	46.64
2=Female	3353836	53.36	6285749	100

Age adjustment

- Age adjustment is a method used to ***compare*** the same health outcome between two populations, or between the same population over time. It is:
  - Used to estimate what the prevalence would be if the age distribution of two populations were the same
  - A relative measure, not an actual measure of risk
- Age-adjusted estimates are only comparable to other age-adjusted estimates that use the same standard population.
- The US 2000 Standard Population is used to age-adjust NYC HANES data, both 2004 and 2013-14.

Weighting vs. age adjustment

- Weighting (**required** to obtain prevalence estimates): Allows the survey to be representative of everyone in NYC population ages 20 and older.
- Age adjustment (**not required**): Facilitates the comparison of
  1. NYC estimates to estimates elsewhere
  2. Estimates between groups within NYC
  3. Estimates in NYC over time

Age adjustment to the US 2000 Standard Population

stdvar AGEGROUP;

stdwgt **0.396579 0.371795 0.231626**;

- The "stdvar" statement designates the standardizing variable. Your standardizing variable must be categorical (here we use age groups: 20-39; 40-59; 60+).
- The "stdwgt" designates the standardizing proportions. Our standardizing proportions are based on the US 2000 Standard Population.
- The stdvar variable must have the same number of categories as the number of stdwgt proportions.

Sampling Design Variables

Stage	2004	2013-14
Stratum	Stratum	Borostratum
Cluster / Nesting	PSU	• PSUnest • HHnest

Step 6: Generate prevalence estimates NYC HANES 2013-14: SUDAAN (1)

```
proc descript data=dbts14_2 filetype=sas design=wr NOTSORTED;
nest Borostratum PSUnest HHnest; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts dbts; /*Variable to be analyzed. State the variable as many times as the categories.*/
catlevel 1 2; /*State the categories of the variable*/
tables _one_ ; /*Prevalence estimate for entire sample*/
subgroup dbts AGEGROUP; /*State all variables in your analysis*/
levels  2 3; /*For each variable in subgroup, state the number of categories*/
stdvar AGEGROUP; /*Age standardization variable*/
stdwgt  0.396579  0.371795  0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle "Age-Standardized Weighted Prevalence of Diabetes, NYC HANES 2013-14 ";
setenv COLWIDTH=16 COLSPCE=2;
print  NSUM="SAMSIZE" WSUM="POPSIZE" total="weights for %"  percent lowpct uppct sepercent / style = nchs nsumfmt=F8.0 wsumfmt=F8.0 totalfmt=F8.0 sepercentfmt=F3.1 percentfmt=F6.1 lowpctfmt=F6.1 uppctfmt=F6.1 ;
run;
```

Step 6: Generate prevalence estimates NYC HANES 2013-14: SUDAAN (2)

	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Diabetes based on FPG, A1c, or Self- Report: 1	1246	6285749	987952	16.1	14	18.4	1.1
Diabetes based on FPG, A1c, or Self- Report: 2	1246	6285749	5297796	83.9	81.6	86.0	1.1

Step 6: Generate prevalence estimates NYC HANES 2013-14: SAS (1)

- In SAS you can use the **proc surveyreg** to get weighted, age-standardized prevalence estimates.
- **proc surveyreg** is an analogous procedure to **proc reg**. Both will give you linear regression estimates.
- Since your outcome variable (dbts) is dichotomous, you need to recode so that a negative outcome is coded as 0 and a positive outcome is coded as 100. That will make it easier to read and interpret the results from the **proc surveyreg**.

```
data dbts14_2;
    set dbts14_2;
    if dbts=2 then dbts_10=0;
    else if dbts=1 then dbts_10=100;
run;
```

Step 6: Generate prevalence estimates NYC HANES 2013-14: SAS (2)

```
proc surveyreg data=dbts14_2 nomcar;
stratum Borostratum; /*Sampling design variable*/
cluster PSUnest HHNest; /*Sampling design variable*/
weight dbts_wt_new; /*Weighting variable*/
class AGEGROUP; /*Age standardization variable*/
model dbts_10=AGEGROUP/noint solution vadjust=none CLPARM;
/*To get prevalence estimate*/
estimate AGEGROUP  0.396579 0.371795 0.231626; /*Standardizing
proportions*/
/*The following concern the appearance of the results*/
TITLE1 "Age-Standardized Weighted Total Prevalence of Diabetes ";
TITLE2 "NYC HANES 2013-14";
run;
```

NYC Health | NYC HANES Health & Nutrition Examination Survey | NYC HANES New York City Health and Nutrition Examination Survey | CUNY SPH

Step 6: Generate prevalence estimates NYC HANES 2013-14: SAS (3)

Estimate								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Row 1	16.0377	1.0884	905	14.73	<.0001	0.05	13.3044	18.1710

STRATIFIED ANALYSIS

Does diabetes prevalence differ by gender?

Stratified Analysis and Significance testing: NYC HANES 2013-14: SUDAAN (1)

1. Gender specific prevalence estimates:

```
proc descript data=dbts14_2 filetype=sas design=wr NOTSORTED;
nest Borostratum PSUnest HHnest; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts; /*Variable to be analyzed*/
catlevel 1; /*State the category of the positive outcome*/
tables gender; /*Prevalence estimate by gender*/
subgroup dbts gender AGEGROUP; /*State all variables in your analysis*/
levels  2 2 3; /*State number of categories for each variable*/
stdvar AGEGROUP; /*Age standardization variable*/
stdwgt 0.396579 0.371795 0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle "Age-Standardized Weighted Prevalence of Diabetes by Gender, NYC HANES 2013-14 ";
setenv COLWIDTH=16 COLSPCE=2;
print  NSUM="SAMSIZE" WSUM="POPSIZE" total="weights for %"  percent lowpct uppct
   sepercent/  style = nchs nsumfmt=F8.0 wsumfmt=F8.0 totalfmt=F8.0 sepercentfmt=F3.1
   percentfmt=F6.1 lowpctfmt=F6.1 uppctfmt=F6.1;
run;
```

Stratified Analysis and Significance testing: NYC HANES 2013-14: SUDAAN (2)

1. Gender specific prevalence estimates:

Diabetes based on FPG, A1c, or Self-Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1246	6285749	986942	16.0	14.0	18.3	1.1
1: Male	523	2931913	449566	15.8	12.9	19.2	1.6
2: Female	723	3353836	537376	16.2	13.4	19.4	1.5

Stratified Analysis and Significance testing: NYC HANES 2013-14: SUDAAN (3)

2. Statistical testing:

```
proc descript data=dbts14_2 filetype=sas design=wr NOTSORTED;
nest Borostratum PSUnest HHNest; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts; /*Variable to be analyzed*/
catlevel 1; /*State the category of the positive outcome*/
class gender AGEGROUP; /*State variables used, except for the “var” variable*/
pairwise gender; /*State variable for statistical testing*/
stdvar AGEGROUP; /*Age standardization variable*/
stdwgt 0.396579 0.371795 0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle “Statistical Test of Diabetes Prevalence by Gender, NYC HANES 2013-14";
setenv DECWIDTH=3;
print NSUM WSUM percent t_pct p_pct / style=nchs nsumfmt=F8.0 wsumfmt=F8.0;
run;
```

Stratified Analysis and Significance testing: NYC HANES 2013-14: SUDAAN (4)

2. Statistical testing:

Diabetes based on FPG, A1c, or Self-Report: 1	Sample Size	Weighted Size	Cntrst Pct	T-Test Cont.Pct=0	P-value T-Test Cont. Pct=0
CONTRAST_1: (1: Male, 2: Female)	1246	6285749	-0.401	-0.180	0.858

Stratified Analysis and Significance testing: NYC HANES 2013-14: SAS (1)

```
proc surveyreg data=dbts14_2 nomcar;
stratum Borostratum; /*Sampling design variable*/
cluster PSUnest HHNest; /*Sampling design variable*/
weight dbts_wt_new; /*Weighting variable*/
class gender AGEGROUP; /*Stratification and age standardization variables*/
model dbts_10 = Gender AGEGROUP gender*AGEGROUP/noint solution vadjust=none
CLPARM; /*Age-standardized prevalence estimate by Gender*/
estimate '1: Male' gender 1 0 AGEGROUP  0.396579 0.371795 0.231626
        gender*AGEGROUP 0.396579 0.371795 0.231626 0 0 0;
estimate '1: Female' gender 0 1 AGEGROUP  0.396579 0.371795 0.231626
        gender*AGEGROUP 0 0 0 0.396579 0.371795 0.231626;
/*The following concern the appearance of the results*/
TITLE1 "Age-Standardized Weighted Total Prevalence of Diabetes by Gender";
TITLE2 "NYC HANES 2013-14";
run;
```

Stratified Analysis and Significance testing: NYC HANES 2013-14: SAS (2)

Estimate								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1: Male	15.8013	1.6194	905	9.76	<.0001	0.05	12.6273	18.9753

Estimate								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1: Female	16.2019	1.4752	905	10.98	<.0001	0.05	13.3105	19.0933

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	6	36.46	<.0001
gender_n	1	0.00	0.9856
agegroup	2	71.24	<.0001
gender_n*agegroup	2	0.31	0.7338

COMPARISON ACROSS SURVEY YEARS

Diabetes Prevalence Estimates

- 2004:

Diabetes based on FPG, A1c, or Self- Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1823	5827719	742365	13.3	11.4	15.6	1.1
1: Male	770	2688231	353753	13.7	11.1	16.9	1.5
2: Female	1053	3139488	388612	13.2	10.7	16.2	1.4

- 2013-14:

Diabetes based on FPG, A1c, or Self- Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1246	6285749	986942	16.0	14.0	18.3	1.1
1: Male	523	2931913	449566	15.8	12.9	19.2	1.6
2: Female	723	3353836	537376	16.2	13.4	19.4	1.5

Before running the analysis (1)

- Appendix: 2004 Analysis provides step-by-step tutorial to adjust weights and get age-standardized estimates of diabetes prevalence for NYC HANES 2004.
- Make sure you are using the same codes and formats both for 2004 and for 2013-14.
- If variable names, codes, or formats are different, recode so that all features match.

Before running the analysis (2)

- Check the variable names and definitions in both survey years
  - Using the NYC HANES 2013-14 codebook, look at the Prior 2004 column. This column indicates the name of the variable in the 2004 survey.

Variable	2004	2013-14
ID	SPID	Key
Diabetes (your recode)	Dbts	Dbts
Gender	riagendr	Gender
Weight variable (your recode)	dbts_wt_new	dbts_wt_new
Sampling Variables	Stratum; PSU; Created HHNest (Slide 73)	Borostratum; PSUnest HHNest

Step 1: Recoding variables

For each survey year:

1. Keep only the variables that will be included in the analysis.
2. Create a year variable.
3. Rename, recode, or reformat those variables, as needed.

```
/*2004 Dataset*/
data dbts_2004;
set dbts04_2;
 rename psu=psunest riagendr=gender stratum=borostratum ageadj=agegroup;
 key=put(sp_id, 7.);
 year=2004;
 keep stratum psu HHNest riagendr dbts key year dbts_wt_new wt_dbts ageadj;
 format riagendr gender. ageadj AGEGROUPFMT. dbts TE_3F. ;
run;
/*2013-14 Dataset*/
data dbts_2014;
set dbts14_2;
 year = 2014;
 keep borostratum psunest HHNest gender dbts key year dbts_wt_new wt_dbts agegroup;
 format dbts TE_3F. ;
run;
```

Step 2: Append the datasets

- Confirm that datasets have same variable names, codes, and formats, and then append the two datasets.

```
/* 1. Verify names and formats*/
proc contents data=dbts_2004; run;
proc contents data=dbts_2014; run;

/* 2. Append the datasets*/
data dbts;
set dbts_2004 dbts_2014;
  /*Restrict dataset by including only participants with a non-missing value for dbts*/
  if dbts=. then delete;
run;
proc freq data=dbts;
  tables year*dbts/list missing;
run;
```

3. Comparison Across Survey Years SUDAAN (1)

```
proc descript data = dbts filetype = sas design = wr notsorted;
nest borostratum psunest HHNest; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts; /*Variable to be analyzed*/
catlevel 1; /*State the category of the positive outcome*/
class  gender agegroup year; /*State all variables in your analysis*/
contrast year = (-1,1) / name = "2004 vs 2013-2014"; /*Compare the two survey
years*/
tables gender; /*Prevalence estimate by gender*/
stdvar agegroup; /*Age standardization variable*/
stdwgt 0.396579  0.371795  0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle "Total Diabetes (age-adjusted and weighted), NYC HANES 2004 vs. 2013-14";
SETENV COLWIDTH=16 COLSPCE=2;
print NSUM="SAMSIZE" percent  sepercent  t_pct  p_pct / style = nchs
nsumfmt=F8.0 percentfmt=F8.3 sepercentfmt=F8.2 p_pctfmt=F8.4;
run ;
```

3. Comparison Across Survey Years SUDAAN (2)

- 2004:

Diabetes based on FPG, A1c, or Self- Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1823	5827719	742365	13.3	11.4	15.6	1.1
1: Male	770	2688231	353753	13.7	11.1	16.9	1.5
2: Female	1053	3139488	388612	13.2	10.7	16.2	1.4

- 2013-14:

Diabetes based on FPG, A1c, or Self-Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1246	6285749	986942	16.0	14.0	18.3	1.1
1: Male	523	2931913	449566	15.8	12.9	19.2	1.6
2: Female	723	3353836	537376	16.2	13.4	19.4	1.5

- Comparison:

2004 vs. 2013-2014	SAMSIZE	Cntrst Pct	SE Cntrst Percent	t-test Cont.Pct=0	p-value t-test Cont.Pct=0
Total	3069	2.698	1.52	1.8	0.0779
Male	1293	2.077	2.27	0.9	0.3618
Female	1776	3.013	2.05	1.5	0.1438

NYC Health | NYC hanes Health & Nutrition Examination Survey | NYC HANES New York City Health and Nutrition Examination Survey | CUNY SPH

SAS vs SUDAAN for Complex Survey Analysis

	Point Estimates (Percents, Means, Etc.)	Variances (Standard Errors, Variances)
Unweighted Regular SAS Procedures	Incorrect	Incorrect
Weighted Regular SAS Procedures	Correct	Incorrect
SAS Proc Survey Procedures (Weighted)	Correct	Correct
SUDAAN Procedures (Weighted)	Correct	Correct

- SAS sometimes has larger p-values and wider confidence intervals than SUDAAN.
- SAS estimates of variance (and thus of standard errors) may be overly conservative, because SAS assumes the first-stage sampling is with replacement, although often it is not.
- SUDAAN allows you to identify with replacement versus without replacement.

Adapted from: http://www2.sas.com/proceedings/sugi31/194-31.pdf

Getting help

- Talk to your colleagues!

- Contacting SUDAAN
  - Go to the technical assistance web page, or
  - Send an email at sudaan@RTI.org

Health

Health & Nutrition Examination Survey

NYC HANES
New York City Health and Nutrition Examination Survey

CUNY SPH

APPENDIX: 2004 ANALYSIS

1. Call in dataset NYC HANES 2004

```
/*Download datasets and save*/
/*For this tutorial, we will assume that datasets are saved in
    C:\Data\2004*/
libname NYCH04 'C:\Data\2004';
options fmtsearch = (formats NYCH14.formats);

/*Each dataset must be sorted on merging variable SP_ID*/
proc sort data = NYCH04.spfile out=spfile; by sp_id;
proc sort data = NYCH04.capi out=capi; by sp_id;
proc sort data = NYCH04.blood out=blood; by sp_id;
data NYCH04;
    merge spfile (in=a) capi blood;
    by sp_id;
    if a; /*Chooses observations that are in the SPfile */
run;
```

Step 2: Identify outcome of interest

1. Identify outcome of interest
   - Diabetes: Diabetes based on FPG, A1c, or self-reported diagnosis:
     - Fasting Plasma Glucose (FPG) equal to or greater than 126 mg/dL, or
     - A1c equal to or greater than 6.5% (regardless of fasting status), or
     - Self-reported diagnosis of diabetes

Step 2: Identify outcome of interest

2. Variables needed for recode

Variable	2004	2013-14
Fasting Indicator	WTSF1F	DBT_Fasting
Glucose	LBXGLU	Glucose
A1c	LBXGH	A1c
Self-Reported Diabetes	DIQ010	DIQ_1

Step 2 continued: Recode/create outcome of interest NYC HANES 2004

```
data dbts04;
set NYCHANES04;

*Define diabetes using FPG or A1c or previous diagnosis of diabetes;

/*Diabetic: Fasted and have plasma glucose >=126 or A1C>=6.5 or self-reported diagnosis*/
if WTSF1F>0 and (lbxglu>=126 or lbxgh>=6.5)  or diq010=1 then dbts=1;

/*Diabetic: Did not fast and have A1c>=6.5 or self-reported diagnosis*/
else if WTSF1F<=0 and lbxgh>=6.5 or diq010=1 then dbts=1;

/*Non-Diabetic: Fasted and have plasma glucose<126 and A1C<6.5 and no self-reported diagnosis*/
else if WTSF1F>0 and (0<lbxglu<126 or 0<lbxgh<6.5)  and diq010 ne 1 then dbts=2;

/* Non-Diabetic:  Did not fast and have A1C<6.5 and no self-reported diagnosis */
else if WTSF1F<=0 and 0<lbxgh<6.5  and diq010 ne 1 then dbts=2;

label dbts = 'Diabetes based on FPG, A1c, or Self-Report';
```

Step 3: Identify appropriate weight (1)

1. Outcome = Diabetes (calculated from blood tests and self-reported diagnosis)
   - Fasting plasma glucose comes from fasting subset
     - Participants who fasted at least 8 hours and have valid glucose laboratory test results
   - A1c comes from blood subset
     - Participants with valid A1c laboratory test results
   - Self-reported diagnosis comes from survey (CAPI)
     - Participants who answered DIQ_1 of the CAPI survey

Step 3: Identify appropriate weight (2)

2. Is our definition of the outcome exclusive or inclusive?
   - If EXCLUSIVE, we define diabetes among ONLY those participants who fasted and who had a valid glucose value.
     - Then use:
       - 2004: Fasting weight (**WTSF1F**)
       - 2013-14: Blood weight (**BLOOD_WT**) and adjust for fasting
     - This definition excludes participants who did not fast as well as participants who did not have a valid glucose measurement, even those with a valid A1c and even those with a diabetes diagnosis.

Step 3: Identify appropriate weight (3)

2. Is our definition of the outcome inclusive or exclusive?
   - If INCLUSIVE, we define diabetes among participants with a valid A1c value OR participants who reported a previous diabetes diagnosis OR participants who both fasted and had a valid plasma glucose value.
     - Then use:
       - 2004: Clinic + Home (**WTSF1CH**)
       - 2013-14: CAPI weight (**CAPI_WT**)

Step 3: Identify appropriate weight (4)

- Number of cases to be included in the analysis: **1823**

```
proc freq data=dbts04;
      tables dbts/list;
Run;
```

Diabetes based on FPG, A1c, or Self-Report				
dbts	Frequency	Percent	Cumulative	Cumulative
1: Yes	191	10.48	191	10.48
2: No	1632	89.52	1823	100
Frequency Missing = 176				

Step 4: Adjust for non-response NYC HANES 2004 (1)

1. Set the initial weight variable of the outcome, the one to be adjusted for non-response

```
data dbts04;
set dbts04;
 /*1. Initial Diabetes weight. This is the first crucial step in adjusting for non-response. Set the weight to zero for cases that will not be included in the analysis (that is, cases with a missing value for the dbts variable).*/
if dbts = . then dbts_Wt=0;
/*2. Set the weight to the identified appropriate weight for participants who will be included in the analysis. (We will use WTSF1CH because our definition is inclusive.)*/
else dbts_Wt = WTSF1CH;
run;
```

Step 4: Adjust for non-response NYC HANES 2004 (2)

2. Create a new dataset called **samplewts**, in which you sum the initial diabetes weight by age, sex, race/ethnicity.
3. Create a new variable called **samplewt** (this will be the sum of the initial diabetes weight for all observations by age, sex, and race/ethnicity):

```
proc summary data = new nway;
 var  dbts_Wt ; /*use the weight created in previous
  step (modified from the clinic+home weight)*/
 class agewt riagendr racewt;
 output out = samplewts sum = samplewt;
run;
```

Step 4: Adjust for non-response NYC HANES 2004 (3)

Partial printout of samplewts

agewt	riagendr	racewt	_TYPE_	_FREQ_	samplewt
20-29	Male	Hispanic	7	76	145570.5
20-29	Male	Non-Hispanic Black	7	36	82361.25
20-29	Male	Non-Hispanic Asian	7	29	60366.44
20-29	Male	Non-Hispanic White/Other	7	86	215247.45
20-29	Female	Hispanic	7	95	146846.8
20-29	Female	Non-Hispanic Black	7	82	163400.72
20-29	Female	Non-Hispanic Asian	7	34	56706.72
20-29	Female	Non-Hispanic White/Other	7	81	183322.25
30-39	Male	Hispanic	7	70	189448.43
30-39	Male	Non-Hispanic Black	7	37	110831.96
30-39	Male	Non-Hispanic Asian	7	28	68934.95
30-39	Male	Non-Hispanic White/Other	7	58	214267.55
30-39	Female	Hispanic	7	115	245985.5
30-39	Female	Non-Hispanic Black	7	48	136758.01
30-39	Female	Non-Hispanic Asian	7	39	83181.36
30-39	Female	Non-Hispanic White/Other	7	58	171544.56

Control totals: NYC HANES 2004

1. These datasets contain reference totals for each of the 3 survey weights
2. The weights in each dataset sum to **5,827,719**

Weight	**Control dataset name**
Clinic + Home (wtsf1ch)	int_con_tots.sas7bdat
Clinic only (wtsf1c)	clin_con_tots.sas7bdat
Fasting (wtsf1f)	fst_con_tots.sas7bdat

NYC Health

NYC hanes Health & Nutrition Examination Survey

NYC HANES New York City Health and Nutrition Examination Survey

CUNY SPH

Step 4: Adjust for non-response NYC HANES 2004 (4)

4. Call in the appropriate control total dataset

```
data int_con_tots;
     set NYCH04.Clinic_home_totals;
run;
```

5. Sort the control total dataset and the samplewts dataset by the variables on which they will merge: agewt riagendr racewt

```
proc sort data=int_con_tots; *clinic + home control totals;
by agewt riagendr racewt;
run;
proc sort data=samplewts; *analysis dataset;
by agewt riagendr racewt;
run;
```

Step 4: Adjust for non-response NYC HANES 2004 (5)

6. Merge the dataset that contains weight control totals with the dataset of summed weights (from the analytic dataset), and create weight factor

```
data wtfactor;
merge samplewts  int_con_tots;
by agewt riagendr racewt;
wt_factor_dbts=clinic_home_control_total/samplewt; /*weight factor*/
keep agewt riagendr racewt wt_factor_dbts;
run;
```

Step 4: Adjust for non-response
NYC HANES 2004 (6)

Partial print out of **wtfactor** dataset

agewt	riagendr	racewt	wt_factor_dbts
20-29	Male	Hispanic	1.03456
20-29	Male	Non-Hispanic Black	1.07284
20-29	Male	Non-Hispanic Asian	1.0983
20-29	Male	Non-Hispanic White/Other	1.14533
20-29	Female	Hispanic	1.04976
20-29	Female	Non-Hispanic Black	1.11994
20-29	Female	Non-Hispanic Asian	1.11162
20-29	Female	Non-Hispanic White/Other	1.10555
30-39	Male	Hispanic	0.99447
30-39	Male	Non-Hispanic Black	1.15524
30-39	Male	Non-Hispanic Asian	1.12221
30-39	Male	Non-Hispanic White/Other	1.08198
30-39	Female	Hispanic	1.05165
30-39	Female	Non-Hispanic Black	1.12494
30-39	Female	Non-Hispanic Asian	1.16688
30-39	Female	Non-Hispanic White/Other	1.08315

Step 4: Adjust for non-response NYC HANES 2004 (7)

7. Sort the new weight factor dataset and your analytic dataset, and then merge the two datasets. Create the weights for your analysis by multiplying the weights of participants with non-missing data by the weight factor.

```
proc sort data=wtfactor; by agewt riagendr racewt; run;
proc sort data=dbts04; by agewt riagendr racewt; run;

data dbts04_2;
        merge dbts04 (in=a) wtfactor;
        by agewt riagendr racewt;
        dbts_wt_new = dbts_Wt * wt_factor_dbts; /*Adjusted Diabetes Weight*/
        if dbts_wt_new = 0 then wt_dbts = 2; /*Adj-diabetes Weight Indicator*/
        else wt_dbts = 1;
        label     dbts_wt_new = 'Diabetes Weight'
        wt_dbts = 'Diabetes Weight Indicator';
        format wt_dbts TE_3F.;
        drop dbts_Wt wt_factor_dbts;
run;
```

Health

NYC hanes Health & Nutrition Examination Survey

NYC HANES New York City Health and Nutrition Examination Survey

CUNY SPH

Step 5: Test your new weight! NYC HANES 2004

1. Make sure your adjusted weight is defined for all the cases included in the analysis: **1823** (see slide 60).

```
proc freq data = dbts04_2;
table wt_dbts;
run;
```

Diabetes Weight Indicator				
wt_dbts	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1: Yes	1823	91.2	1823	91.2
2: No	176	8.8	1999	100

2. Make sure total weighted frequency is the total population represented by NYC HANES, **5,827,719**, and the gender distribution matches that of the survey, 46.1% male and 53.9% females:

```
proc freq data = dbts04_2;
table riagendr ;
weight dbts_wt_new;
run;
```

Gender				
riagendr	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	2688231	46.13	2688231	46.13
Female	3139488	53.87	5827719	100

Age adjustment

- Age adjustment is a method used to ***compare*** the same health outcome between two populations, or between the same population over time. It is:
  - Used to estimate what the prevalence would be if the age distribution of two populations were the same
  - A relative measure, not an actual measure of risk
- Age-adjusted estimates are only comparable to other age-adjusted estimates that use the same standard population.
- The US 2000 Standard Population is used to adjust NYC HANES data, both 2004 and 2013-14.

Weighting vs. age adjustment

- Weighting (**required** to obtain prevalence estimates): allows the survey sample to be representative of everyone in NYC
- Age adjustment (**not required**): facilitates the comparison of
  1. NYC estimates to estimates elsewhere
  2. Estimates between groups within NYC
  3. Estimates in NYC over time

Age adjustment to the US 2000 Standard Population

stdvar ageadj;

stdwgt **0.396579 0.371795 0.231626**;

- The "stdvar" statement designates the standardizing variable. Your standardizing variable must be categorical (here we use age groups = 20-39; 40-59; 60+).
- The "stdwgt" designates the standardizing proportions. Our standardizing proportions are based on the US 2000 Standard Population.
- The stdvar variable must have the same number of categories as the number of stdwgt proportions.

Sampling Design Variables

Stage	2004	2013-14
Stratum	Stratum	Borostratum
Cluster / Nesting	PSU	• PSUnest • HHnest

Please note:

- **In 2004 we did not have a household level clustering variable (HHNest in 2013-14) as part of the sampling design variables.**
- **For comparing estimates across survey years we will create such a variable.**

```
/*  Create HHNest variable in 2004 dataset    */
data dbts04_2;
set dbts04_2;
HHNest = _n_; /*This ensures the HHNest has a unique value for each case*/
run;
```

Step 6: Generate prevalence estimates NYC HANES 2004: SUDAAN (1)

```
proc descript data=dbts04_2 filetype=sas design=wr NOTSORTED;
nest stratum psu; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts dbts; /* Variable to be analyzed. State the variable as many times as the categories*/
catlevel 1 2; /*State the categories of the variable*/
tables _one_; /*Prevalence estimate for entire sample*/
subgroup dbts ageadj; /*State all variables in your analysis*/
levels  2 3; /* For each variable in subgroup, state the number of categories*/
stdvar ageadj; /*Age standardization variable*/
stdwgt 0.396579 0.371795 0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle "Age-Standardized Weighted Prevalence of Diabetes, NYC HANES 2004";
setenv COLWIDTH=16 COLSPCE=2;
print  NSUM="SAMSIZE" WSUM="POPSIZE" total="weights for %"  percent lowpct uppct
sepercent/  style = nchs nsumfmt=F8.0 wsumfmt=F8.0 totalfmt=F8.0 sepercentfmt=F3.1
percentfmt=F6.1 lowpctfmt=F6.1 uppctfmt=F6.1;
run;
```

Step 6: Generate prevalence estimates NYC HANES 2004: SUDAAN (2)

	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Diabetes based on FPG, A1c, or Self-Report: 1	1823	5827719	742365	13.3	11.4	15.6	1.1
Diabetes based on FPG, A1c, or Self-Report: 2	1823	5827719	5085354	86.7	84.4	88.6	1.1

Step 6: Generate prevalence estimates NYC HANES 2004: SAS (1)

- In SAS you can use the **proc surveyreg** to get weighted, age-standardized prevalence estimates.
- **proc surveyreg** is an analogous procedure to **proc reg**. Both will give you linear regression estimates.
- Since your outcome variable (dbts) is dichotomous, you need to recode so that a negative outcome is coded as 0 and a positive outcome is coded as 100. That will make it easier to read and interpret the results from the **proc surveyreg**.

```
/*Recode dbts*/
data dbts04_2;
    set dbts04_2;
    if dbts=2 then dbts_10=0;
    else if dbts=1 then dbts_10=100;
run;
```

Step 6: Generate prevalence estimates NYC HANES 2004: SAS (2)

```
proc surveyreg data=dbts04_2 nomcar;
stratum stratum ; /*Sampling design variable*/
cluster PSU; /*Sampling design variable*/
weight dbts_wt_new; /*Weighting variable*/
class ageadj; /*Age standardization variable*/
model dbts_10=ageadj/noint solution vadjust=none CLPARM;
/*To get age standardized prevalence estimate*/
estimate ageadj  0.396579 0.371795 0.231626;/*Standardizing
proportions*/
/*The following concern the appearance of the results*/
TITLE1 "Age-Standardized Weighted Total Prevalence of Diabetes";
TITLE2 "NYC HANES 2004";
run;
```

Step 6: Generate prevalence estimates NYC HANES 2004: SAS (3)

Estimate								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Row 1	13.3392	1.0634	143	12.54	<.0001	0.05	11.2371	15.4412

STRATIFIED ANALYSIS

How different is the diabetes prevalence by gender?

Stratified Analysis and Significance testing: NYC HANES 2004: SUDAAN (1)

1. Gender specific prevalence estimates:

```
proc descript data=dbts04_2 filetype=sas design=wr NOTSORTED;
nest stratum psu; /*Sampling design variables*/
weight dbts_wt_new; /*Weight*/
var dbts; /*Variable to be analyzed.*/
catlevel 1; /*State the categories of the positive outcome*/
tables riagendr; /*Prevalence estimate by gender*/
subgroup dbts riagendr ageadj; /*State all variables in your analysis*/
levels  2 2 3; /* For each variable in subgroup, state the number of categories*/
stdvar ageadj; /*Age standardization variable*/
stdwgt 0.396579 0.371795 0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
rtitle "Age-Standardized Weighted Prevalence of Diabetes, NYC HANES 2004";
setenv COLWIDTH=16 COLSPCE=2;
print  NSUM="SAMSIZE" WSUM="POPSIZE" total="weights for %"  percent lowpct uppct sepercent/  style =
nchs  nsumfmt=F8.0 wsumfmt=F8.0 totalfmt=F8.0 sepercentfmt=F3.1 percentfmt=F6.1 lowpctfmt=F6.1
uppctfmt=F6.1;
run;
```

Stratified Analysis and Significance testing: NYC HANES 2004: SUDAAN (2)

1. Gender specific prevalence estimates:

Diabetes based on FPG, A1c, or Self-Report: 1	SAMSIZE	POPSIZE	Weights for %	Percent	Lower 95% Limit Percent	Upper 95% Limit Percent	SE Percent
Total	1823	5827719	742365	13.3	11.4	15.6	1.1
Male	770	2688231	353753	13.7	11.1	16.9	1.5
Female	1053	3139488	388612	13.2	10.7	16.2	1.4

Stratified Analysis and Significance testing: NYC HANES 2004: SUDAAN (3)

2. Statistical testing:

```
proc descript data=dbts04_2 filetype=sas design=wr NOTSORTED;
 nest stratum psu; /*Sampling design variables*/
 weight dbts_wt_new; /*Weight*/
 var dbts; /*Variable to be analyzed*/
 catlevel 1; /*State the category of the positive outcome*/
 class  riagendr ageadj; /*State variables used, except for the "var" variable*/
 pairwise riagendr; /*State variable for statistical testing*/
 stdvar ageadj; /*Age standardization variable*/
 stdwgt 0.396579 0.371795 0.231626; /*Standardizing proportions*/
/*The following concern the appearance of the results*/
 rtitle "Statistical Test of Diabetes Prevalence by Gender, NYC HANES 2004";
 setenv DECWIDTH=3;
 print    NSUM WSUM percent  sepercent  t_pct  p_pct /
    style=nchs  nsumfmt=F8.0  wsumfmt=F8.0 ;
run;
```

Stratified Analysis and Significance testing: NYC HANES 2004: SUDAAN (4)

2. Statistical testing:

Diabetes based on FPG, A1c, or Self-Report: 1	Sample Size	Weighted Size	Cntrst Pct	SE Cntrst Pct	T-Test Cont.Pct=0	P-value T-Test Cont.Pct=0
CONTRAST_1: (Male, Female)	1823	5827719	0.536	1.953	0.274	0.784

Stratified Analysis and Significance testing: NYC HANES 2004: SAS (1)

```
proc surveyreg data=dbts04_2 nomcar;
stratum stratum ; /*Sampling design variable*/
cluster PSU; /*Sampling design variable*/
weight dbts_wt_new; /*Weighting variable*/
class   riagendr ageadj; /*Stratification and age standardization variables*/
model dbts_10 = riagendr ageadj riagendr*ageadj/noint solution vadjust=none CLPARM;
/*To get age standardized prevalence estimate by Gender*/
estimate '1: Male' riagendr 1 0 ageadj  0.396579 0.371795 0.231626
          riagendr*ageadj 0.396579 0.371795 0.231626 0 0 0;
estimate '2: Female' riagendr 0 1 ageadj  0.396579 0.371795 0.231626
          riagendr*ageadj 0 0 0 0.396579 0.371795 0.231626;
/*The following concern the appearance of the results*/
TITLE1 "Age-Standardized Weighted Total Prevalence of Diabetes by Gender";
TITLE2 "NYC HANES 2004";
run;
```

Stratified Analysis and Significance testing: NYC HANES 2004: SAS (2)

Estimate								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
1: Male	13.7241	1.4792	143	9.28	<.0001	0.05	10.8003	16.648

Estimate								
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
2: Female	13.1885	1.3965	143	9.44	<.0001	0.05	10.428	15.949

Tests of Model Effects			
Effect	**Num DF**	**F Value**	**Pr > F**
Model	6	29.52	<.0001
riagendr	1	0	0.9823
ageadj	2	58.7	<.0001
riagendr*ageadj	2	1.05	0.3524