

Program Summary - code.sas

Execution Environment

Author: u62298616
File: /home/u62298616/HDAT9400/ASSI2B/code.sas
SAS Platform: Linux LIN X64 3.10.0-1062.12.1.el7.x86_64
SAS Host: ODAWS01-APSE1-2.ODA.SAS.COM
SAS Version: 9.04.01M7P08062020
SAS Locale: en_GB
Submission Time: 21/02/2024, 20:48:19
Browser Host: 220.235.188.242
User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/121.0.0.0 Safari/537.36 Edg/121.0.0.0
Application Server: ODAMID00-APSE1-2.ODA.SAS.COM

Code: code.sas

```
*****
* Project: HDAT9400 - Management and curation of health data
* Purpose: ASSIGNMENT 2B - Creating Analysis-ready data for reproducible research
* Inputs: assignmt2b_ed_data.sas7bdat assignmt2b_formats.sas assignmt2b_gp_data.sas7bdat assignmt2b_pbs_data.sas7bdat
* Location: HDAT9400/ASSI2B
* Author: Hannah Mun
* Date: 23 NOV 2022
*****



*****  
/* SETUP */  
*****  
* Set the root location for the project folder;  
%let path = /home/u62298616/HDAT9400/ASSI2B;  
  
* Include a macro program to run formats script;  
%include "&path/assignmt2b_formats.sas";  
  
* Create a new library;  
Libname assi2b "&path";  
  
* Tell SAS to ignore missing formats;  
options nofmterr;  
  
*****  
/* 0. Exploratory data analysis */  
*****  
* Import the data to the temporary WORK library;  
data gp; set assi2b.assignmt2b_gp_data; run;  
data ed; set assi2b.assignmt2b_ed_data; run;  
data pbs; set assi2b.assignmt2b_pbs_data; run;  
  
***** 0.1 GP data *****  
  
* Check contents of GP dataset;  
proc contents data=gp order=varnum; run;      * 5300 records and 18 variables are observed;  
  
* View the first 10 rows of data;  
proc print data=gp (obs=10);  
run;  
  
* Descriptive statistics of GP dataset;  
proc means data=gp mean median mode std var min max nmiss maxdec=2;  
var ID GP_last age drinks_day height weight syst_bp diast_bp BMI_GP;  
run;  
  
proc freq data=gp;  
table sex cob healthcare_card adverse_reaction Smoke_current_GP Risky_alcohol_GP Obese_GP HighBP_GP reason/nocum;  
run;  
  
* Univariate analysis;  
ods graphics/ height=500px width=500px;  
proc univariate data=gp noprint;  
hist age drinks_day height weight syst_bp diast_bp BMI_GP /normal;
```

```

run;

* Bivariate Analysis;
proc sgpanel data=gp;
  panelby sex;
  vbar HighBP_GP /
    response=Obese_GP
    stat=mean;
run;

***** 0.2 ED data *****

* Check contents of dataset;
proc contents data=ed order=varnum; run;      * 63614 records and 15 variables are observed;

* View the first 10 rows of data;
proc print data=ed (obs=10);
run;

* Descriptive statistics of dataset and check missing values for numeric variables;
proc means data=ed mean median mode std var min max nmiss maxdec=2;
var ID age_ed ed_admission ed_separation;
run;

* Descriptive statistics of dataset and check missing values for categorial variables;
proc freq data=ed;
table sex_ed cob_ed interpreter health_insurance triage_category separation_mode/nocum ;
run;

* Check missing values in dx1-dx5;
proc sql;
select nmiss(dx1) as dx1,nmiss(dx2) as dx2,nmiss(dx3) as dx3,nmiss(dx4) as dx4,nmiss(dx5) as dx5
from ed;
quit;

* Univariate analysis;
ods graphics on;
proc freq data=ed ;
tables sex_ed cob_ed interpreter health_insurance triage_category separation_mode/ plots=freqplot;
run;

proc univariate data=ed noprint;
hist age_ed/normal;
run;

***** 0.3 PBS data *****

* Check contents of dataset;
proc contents data=pbs order=varnum; run;      * 3164 records and 6 variables are observed;

* View the first 10 rows of data;
proc print data=pbs (obs=10);
run;

* Descriptive statistics of dataset and check missing values for numeric variables;
proc means data=pbs mean median mode std var min max nmiss maxdec=2;
var ID supply_date;
run;

* Find unique values in character type variables;
proc sql;
select count(distinct 'ATC'n) as 'ATC'n,
       count(distinct 'drug_name'n) as 'drug_name'n,
       count(distinct 'item_code'n) as 'item_code'n,
       count(distinct 'form_strength'n) as 'form_strength'n
from pbs;
quit;

* Check missing value in character type variables;
proc sql;
select nmiss(ATC) as ATC, nmiss(drug_name) as drug_name, nmiss(item_code) as item_code,nmiss(form_strength) as form_strength
from pbs;
quit;

* Univariate analysis;
ods graphics on;
proc freq data=ed ;

```

```

tables sex_ed cob_ed interpreter health_insurance triage_category separation_mode/ plots=freqplot;
run;

proc univariate data=ed nopolish;
hist age_ed/normal;
run;

*****
/* 1. Research Question 1 (Primary care perspective) */
*****;

/* 1.1. Create a new variable Agegroup_GP */

* Create a format for new variable;
proc format;
  value agegroup
    1='Under 60 years old' 2='60 years old and older';
run;
* Create a new variable and apply format;
data gp1;
  set gp;
  * AgegroupGP =1 : Under 60 years old, AgegroupGP =2 : 60 years old and older;
  if age < 60 then Agegroup_GP =1;
  if age =>60 then Agegroup_GP =2;
  format Agegroup_GP agegroup.;
run;
* Check new variable created;
proc freq data=gp1; table Agegroup_GP /missing;
run;

/* 1.2. Calculate the proportion of GP patients who attended the ED in 2014 and interpret the finding */

* Merge gp1 and ed data and add a new variable;
data merge1;
  merge gp1(in=a) ed(in=b); by ID;
  if b then ed_record=1;
  else ed_record=0;
  format ed_record ynf.;
  if a;
run;

* Calculate the proportion of GP patients who attended ED in 2014;
proc sort data=merge1 out=merge1_nodup nodupkey; * output dataset is excluded duplicated ID;
by id ed_record;

proc freq data=merge1_nodup;
  table ed_record;
run;

/* 1.3. Calculate total number of monthly ED admissions for all GP patients. Plot the result to show monthly trends and interpret */

* Create a new variable month that contains month of ed_admission date;
data merge2;
set merge1_nodup;
month = month(ed_admission);
run;

* Calculate total number of monthly ED admissions for all GP patients;
proc sql;
select count(*) as Total_number from merge2 where ed_record=1;
select count(*) as Total_number from merge2 where ed_record=0; * no. of GP patients who don't have ED admission record for data;
quit;

* Count number of ed_admission date of all GP patients by monthly;
proc sql;
create table monthly as
select month, count(*) as MonAdm from merge2
group by month;
quit;

* Plot monthly ed_admission ;
title 'Distribution of monthly ED admission for all GP patients';
proc sgplot data=monthly;
vbar month / response=MonAdm ;

```

```

xaxis display=(nolabel)
  values=(1 2 3 4 5 6 7 8 9 10 11 12)
  valuesdisplay=('Jan' 'Feb' 'Mar' 'Apr' 'May' 'Jun' 'Jul' 'Aug' 'Sep' 'Oct' 'Nov' 'Dec');
yaxis grid label='ED admission';
run;

/* 1.4. Examine differences between patients who did and did not attend the ED in 2014 */
/* in terms of socio-demographic characteristics [sex, age group, country of birth, and health care card] */
/* and health-related factors [smoking, risky alcohol consumption, obesity, and high blood pressure] */

* Compare socio-demographic characteristics and health-related factors by ed_record;
proc freq data=merge2;
tables (sex agegroup_gp cob healthcare_card smoke_current_gp risky_alcohol_gp obese_gp highbp_gp) * ed_record /chisq norow noo
run;

/* 1.5. Among GP patients who visited the ED, calculate the total number of ED attendance for each person in 2004. */
/* Describe the distribution of numbers of ED attendance using a histogram and descriptive statistics. */

* GP patients who has ED admission record;
data attendance; set merge1;
where ed_record eq 1;
run;

* Count number of ED attendance per each ID;
proc sort data=attendance ;
by id; run;

data att_count;
set attendance;
by id;
  * Generate admseq - sequence within each person ID;
  retain admseq;
  if first.id then admseq=1;
  else admseq = admseq + 1;
  * Produce output with the last number of admseq of each ID;
  if last.id then output;
run;

* Describe distribution of number of ED admission;
proc means data=att_count (keep= id admseq) min median mean max q1 q3 p10 p25 p50 p75 p90 p99;
run;
* Plot the distribution;
title 'Distribution of number of ED admission per single GP patient';
proc univariate data=att_count noprint;
histogram admseq / normal;
inset N = 'Number of GP patients' Mean(8.2) Median(8.2) STD = 'Standard Deviation' (8.3) / position=ne;
label admseq='Number of ED adimission' ;
run;

/* 1.6. Continue with the results of step 1.5, select GP patients who had a large number of ED visits (i.e. top quartile).
Examine and report socio-demographic and health-related characteristics of these patients */

* Select GP patients in top quartile (top 25 % of ED admission counts);
data top_count; set att_count;
where admseq > 6; /*75% quantile in acsending order is top 25% in descending order;
run;

title 'Socio-demographic report of the highest 25% of ED visit patients';
proc freq data=top_count;
tables sex agegroup_gp cob healthcare_card /plots=freqplot ;
run;

title 'Health-related characteristics report of the highest 25% of ED visit patients';
proc freq data=top_count;
tables smoke_current_gp risky_alcohol_gp obese_gp highbp_gp/plots=freqplot ;
run;

*****/* 2. Research Question 2 (ED care perspective) */
*****;

/* 2.1. Create three variables to flag ED records with these behaviours being recorded in any diagnosis field */

* Create a format of new variables;
proc format;
  value smoker
    0='No' 1='Yes, smoker';

```

```

value drinker
0='No' 1='Yes, drinker';
value obese
0='No' 1='Yes, obese';
run;

/* Create new variables;
data ed1; set ed;
array diag {*} dx1-dx5;
smoker_flag=0;
risky_alcohol_flag=0;
obesity_flag=0;
do i=1 to dim(diag);
  if diag{i} in ('F17','Z72') then smoker_flag=1; * smoker_flag = 0 : No, smoker_flag = 1 : Yes, smoker;
  if diag{i} = 'F10' then risky_alcohol_flag=1;   * risky_alcohol_flag = 0 : No, risky_alcohol_flag = 1 : Yes, drinker;
  if diag{i} = 'E66' then obesity_flag=1;         * obesity_flag = 0 : No, obesity_flag = 1 : Yes, obese;
drop i;
end;
format smoker_flag smoker.;
format risky_alcohol_flag drinker.;
format obesity_flag obese.;
run;

/* 2.2. Classify whether the patient smokes, drinks alcohol at risky level or is obese,
if these risk factors are recorded in any ED record for a patient.
Calculate and report the prevalence of smoking, risky alcohol consumption and obesity among ED patients */

/* Create new variables;
data ed2; set ed1;
  * Assign default value of new variables;
smoker_ED=0;
risky_alcohol_ED=0;
obesity_ED=0;
  * Apply condition;
if smoker_flag = 1 then smoker_ED =1;
if risky_alcohol_flag =1 then risky_alcohol_ED =1;
if obesity_flag =1 then obesity_ED =1;
* Apply format;
format smoker_ED smoker.;
format risky_alcohol_ED drinker.;
format obesity_ED obese.;
run;

* Summarize dataset per each ED patient ID;
proc sort data=ed2; by id; run;
proc means noprint data=ed2; by id; output out=ed3 (drop = _type_ _freq_)
  max(smoker_ED) =smoker_ED
  max(risky_alcohol_ED) = risky_alcohol_ED
  max(obesity_ED) = obesity_ED;
run;

* Calculate prevalence of smoking, risky alcohol consumption and obesity;
proc freq data=ed3;
tables smoker_ED risky_alcohol_ED obesity_ED ;
run;

/* 2.3. Calculate and report the proportion of ED patients who had a visit to Medical Plus GP in 2014 */
* Merge ed2 and gp data and add a new variable gp_record;
data merge3;
  merge ed2(in=a) gp(in=b); by id;
  if b then gp_record=1;
  else gp_record=0;
  format gp_record ynf.;
  if a;
run;

* Calculate proportion of ED patients who has gp_record or not;
proc sort data=merge3 out=merge3_nodup nodupkey;
by id gp_record;
proc freq data=merge3_nodup;
table gp_record;
run;

/* 2.4. Examine whether there are any differences between ED patients who did and did not visit a GP  */

```

```

/* in terms of sex, age, country of birth, private health insurance, smoking, risky alcohol consumption and obesity. */
/* You can categorise patient age into two groups (under 60 /60 and older). */

* Create a new variable agegroup_ed and apply format;
data merge4;
  set merge3_nodup;
  * AgegroupGP =1 : Under 60 years old, AgegroupGP =2 : 60 years old and older;
  if 0< age_ed < 60 then agegroup_ed =1;
  if age_ed =>60 then agegroup_ed =2;
  format agegroup_ed agegroup.;
run;

proc freq data=merge4;
tables (sex_ed agegroup_ed cob_ed health_insurance smoker_ed risky_alcohol_ed obesity_ed) * gp_record /chisq norow nopercent
run;

* Plot data;
proc sgpanel data=merge4;
panelby gp_record;
vbar agegroup_ed ;
run;

proc sgpanel data=merge4;
panelby gp_record;
vbar cob_ed ;
run;

proc sgpanel data=merge4;
panelby gp_record;
vbar smoker_ed ;
run;

proc sgpanel data=merge4;
panelby gp_record;
vbar risky_alcohol_ed ;
run;

proc sgpanel data=merge4;
panelby gp_record;
vbar obesity_ed ;
run;

/* 2.5. Calculate overall sensitivity (Sn) and specificity (Sp) of the recording of patient smoking in the ED data, */
/* using patient smoking information in the GP data as the gold standard. */
/* Comment on overall quality of ED data on patient smoking */

* Create new variables - TP(true positive),TN(true negative),FP(false positive) and FN(false negative);
data sn_sp; set merge4;
  if smoke_current_gp =1 then do;
    if smoker_ed =1 then result ="TP";
    else if smoker_ed =0 then result = "FN";
  end;
  else if smoke_current_gp =0 then do;
    if smoker_ed = 1 then result = "FP";
    else if smoker_ed = 0 then result = "TN";
  end;
run;

* Count number of TP, FN, FP, TN;
proc sort data=sn_sp;
  by smoke_current_gp smoker_ed;
run;

data sn_sp2 (keep= tp fn fp tn);
  set sn_sp;
  by smoke_current_gp;
  retain tp fn fp tn;
  if first.smoke_current_gp then do;
    tp=0; fn=0; fp=0; tn=0;
  end;
  if result in ("TP") then tp=tp+1;
  if result in ("FN") then fn=fn+1;
  if result in ("FP") then fp=fp+1;
  if result in ("TN") then tn=tn+1;
  else;
  if last.smoke_current_gp then output;
run;

```

```

run;

* Create a table of counted elements;
proc sql;
  create table sn_sp3 as
    select sum(tp) as TP, sum(fn) as FN, sum(fp) as FP, sum(tn) as TN
    from sn_sp2;
quit;

* Calculate sensitivity and specificity;
proc sql;
  create table sn_sp4 as
    select tp/(tp+fn) as Sensitivity, tn/(tn+fp) as Specificity, (tn+tp)/(tn+tp+fn+fp) as Accuracy
    from sn_sp3;
quit;

/* 2.6. Repeat calculation of Sn and Sp of the recording of smoking in ED data, */
/* separately for each patient's sex, age group, country of birth and private health insurance (i.e. stratified by sociodemog */
/* Comment on whether recording of smoking information in ED data differs by patient sociodemographic characteristics */

* Calculate Sn, Sp by sex;
proc sort data=sn_sp; by sex_ed; run;

proc freq data=sn_sp;
by sex_ed;
tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
run;

* Calculate Sn, Sp by age group;
proc sort data=sn_sp; by agegroup_ed; run;

proc freq data=sn_sp;
by agegroup_ed;
tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
run;

* Calculate Sn, Sp by country of birth;
proc sort data=sn_sp; by cob_ed; run;

proc freq data=sn_sp;
by cob_ed;
tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
run;

* Calculate Sn, Sp by private health insurance;
proc sort data=sn_sp; by health_insurance; run;

proc freq data=sn_sp;
by health_insurance;
tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
run;

** Note: in above calculation using 'senspec' reference cell is set TN instead of TP,
therefore sensitivity and specificity value in result table are displayed in opposite way.
I tried to apply table option 'SENSPEC REFCELL=' but could not produce the result by getting syntax error;

***** ****
/* 3. Research Question 3 (Tobacco control perspective) */
***** *****;

/* 3.1. Create a cohort of Sunnydale residents who smoke using information from the GP and ED data sources. */
/* How many smokers could you identify in the GP data alone, ED data alone, and a combination of both GP/ED data sources */

* Identify unique ID without duplicated in GP and ED data;
proc sort data=gp1 out=gp_id nodupkey; by id; run;
proc sort data=ed2 out=ed_id nodupkey; by id; run;

* Merge GP and Ed;
data merge5 ;
merge gp_id(in=a) ed_id(in=b);
if a or b;
by id;
run;

* Check if there are duplicated id;
proc sort data=merge5 out=dup_check nodupkey; by id; run;

```

```

* Count number of smokers identified in each dataset;
proc sql;
select count(id) as gp_smoker   from gp_id where smoke_current_gp eq 1;
select count(id) as ed_smoker   from ed_id where smoker_ed eq 1;
select count(id) as gped_smoker from merge5 where (smoke_current_gp eq 1) or (smoker_ed eq 1);
quit;

* Cleanup merged dataset and crete cohort group;
proc sort data=merge5(drop= sex age age_ed cob dx1 dx2 dx3 dx4 dx5 smoker_flag risky_alcohol_flag obesity_flag) out=cohort no
by id;
where (smoke_current_gp eq 1) or (smoker_ed eq 1); * cohort group include only smokers;
run;

* remove complete duplicates;
* keep sex_ed and drop sex (sex matches for all ID but sex_ed has more data compared to sex);
* keep cob_ed and drop cob
* drop age and age_ed and keep agegroup_gp (ther is a slight different in age and age_ed of some ID, so just using agegroup)
* drop dx1-dx5
* drop smoker_flag risky_alcohol_flag obesity_flag

* Check cohort dataset;
proc contents data=cohort varnum; run;

* See the first 10 rows of the cohort data;
proc print data=cohort (obs=10); run;

/* 3.2. Examine PBS data against the cohort defined in Step 3.1 and comment on the value of PBS data as an additional data sou
/* to identify people who smoke and who were not identified in GP or ED data */

* Classify ID in PBS data who has PBS record of N07BA01, N06AX12 or N07BA03;
data pbs2(keep=id atc); set pbs;
where ATC = 'N07BA01' or ATC = 'N06AX12' or ATC = 'N07BA03';
run;

proc freq data=pbs2;
tables id atc /missing;
run;

* Find IDs in PBS data that is not in cohort(GP&ED) data;
proc sql;
create table missing_smoker as
select id
from pbs2 cohort
where pbs2.id not in (select id from cohort);
quit;

/* 3.3. For the cohort created in Step 3.1, calculate the proportion of smokers who used any of the smoking cessation therapie
/* as well as each of the three individual medicine in 2014. */

* Merge cohort and PBS data;
data merge6;
merge cohort(in=a) pbs2(in=b);
if a;
by id;
run;

* Calculate proportion who used any of the smoking cessation therapies (where ATC is NULL/ATC is NOT NULL);
proc freq data=merge6;
  table atc /plots=freqplot;
...

```

Log: code.sas

Notes (220)

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
69
70      ****
71      * Project: HDAT9400 - Management and curation of health data
72      * Purpose: ASSIGNMENT 2B - Creating Analysis-ready data for reproducible research
73      * Inputs: assignmt2b_ed_data.sas7bdat assignmt2b_formats.sas assignmt2b_gp_data.sas7bdat assignmt2b_pbs_data.sas7bdat
74      * Location: HDAT9400/ASSI2B
75      * Author:Hannah Mun
76      * Date:23 NOV 2022
77      ****

```

```

78
79
80 ****
81 /* SETUP */
82 ****
83 * Set the root location for the project folder;
84 %let path = /home/u62298616/HDAT9400/ASSI2B;
85
86 * Include a macro program to run formats script;
87 %include "&path/assignmt2b_formats.sas";
NOTE: Format SEXF has been output.
NOTE: Format COBF has been output.
NOTE: Format YNF has been output.
NOTE: Format TRIAGEF has been output.
NOTE: Format SEPMODEF has been output.

NOTE: PROCEDURE FORMAT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.01 seconds
      system cpu time 0.00 seconds
      memory        288.93k
      OS Memory     19360.00k
      Timestamp     21/02/2024 09:48:09 AM
      Step Count    24  Switch Count  2
      Page Faults   0
      Page Reclaims 86
      Page Swaps    0
      Voluntary Context Switches 13
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 56

113
114 * Create a new library;
115 Libname assi2b "&path";
NOTE: Libref ASSI2B was successfully assigned as follows:
      Engine: V9
      Physical Name: /home/u62298616/HDAT9400/ASSI2B
116
117 * Tell SAS to ignore missing formats;
118 options nofmterr;
119
120
121 ****
122 /* 0. Exploratory data analysis */
123 ****
124 * Import the data to the temporary WORK library;
125 data gp; set assi2b.assignmt2b_gp_data; run;
NOTE: Data file ASSI2B.ASSIGNMT2B_GP_DATA.DATA is in a format that is native to another host, or the file encoding does not match
      the session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might
      reduce performance.

NOTE: There were 5300 observations read from the data set ASSI2B.ASSIGNMT2B_GP_DATA.
NOTE: The data set WORK.GP has 5300 observations and 18 variables.
NOTE: DATA statement used (Total process time):
      real time      0.01 seconds
      user cpu time  0.01 seconds
      system cpu time 0.00 seconds
      memory        2555.21k
      OS Memory     22440.00k
      Timestamp     21/02/2024 09:48:09 AM
      Step Count    25  Switch Count  3
      Page Faults   0
      Page Reclaims 689
      Page Swaps    0
      Voluntary Context Switches 25
      Involuntary Context Switches 0
      Block Input Operations 1696
      Block Output Operations 1800

126 data ed; set assi2b.assignmt2b_ed_data; run;
NOTE: Data file ASSI2B.ASSIGNMT2B_ED_DATA.DATA is in a format that is native to another host, or the file encoding does not match
      the session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might
      reduce performance.

NOTE: There were 30466 observations read from the data set ASSI2B.ASSIGNMT2B_ED_DATA.
NOTE: The data set WORK.ED has 30466 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time      0.03 seconds
      user cpu time  0.03 seconds
      system cpu time 0.01 seconds
      memory        3043.21k
      OS Memory     23464.00k
      Timestamp     21/02/2024 09:48:09 AM

```

```

Step Count          26  Switch Count  2
Page Faults        0
Page Reclaims      358
Page Swaps         0
Voluntary Context Switches  21
Involuntary Context Switches 1
Block Input Operations 5792
Block Output Operations 5896

127      data pbs; set assi2b.assignmt2b_pbs_data; run;
NOTE: Data file ASSI2B.ASSIGNMT2B_PBS_DATA is in a format that is native to another host, or the file encoding does not match
the session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might
reduce performance.

NOTE: There were 3164 observations read from the data set ASSI2B.ASSIGNMT2B_PBS_DATA.
NOTE: The data set WORK.PBS has 3164 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time      0.01 seconds
      user cpu time  0.01 seconds
      system cpu time 0.00 seconds
      memory       3036.46k
      OS Memory    23464.00k
      Timestamp    21/02/2024 09:48:09 AM
      Step Count     27  Switch Count  2
      Page Faults    0
      Page Reclaims   365
      Page Swaps      0
      Voluntary Context Switches 22
      Involuntary Context Switches 0
      Block Input Operations 4384
      Block Output Operations 4360

128
129      ***** 0.1 GP data *****
130
131      * Check contents of GP dataset;
132      proc contents data=gp order=varnum; run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time      0.03 seconds
      user cpu time  0.04 seconds
      system cpu time 0.00 seconds
      memory       2853.90k
      OS Memory    24104.00k
      Timestamp    21/02/2024 09:48:09 AM
      Step Count     28  Switch Count  0
      Page Faults    0
      Page Reclaims   969
      Page Swaps      0
      Voluntary Context Switches 2
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 32

132      !
                           * 5300 records and 18 variables are observed;

133
134      * View the first 10 rows of data;
135      proc print data=gp (obs=10);
136      run;

NOTE: There were 10 observations read from the data set WORK.GP.
NOTE: PROCEDURE PRINT used (Total process time):
      real time      0.02 seconds
      user cpu time  0.03 seconds
      system cpu time 0.00 seconds
      memory       1588.71k
      OS Memory    23844.00k
      Timestamp    21/02/2024 09:48:09 AM
      Step Count     29  Switch Count  0
      Page Faults    0
      Page Reclaims   325
      Page Swaps      0
      Voluntary Context Switches 0
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 8

137
138      * Descriptive statistics of GP dataset;
139      proc means data=gp mean median mode std var min max nmiss maxdec=2;
140      var ID GP_last age drinks_day height weight syst_bp diast_bp BMI_GP;
141      run;

```

NOTE: There were 5300 observations read from the data set WORK.GP.

NOTE: PROCEDURE MEANS used (Total process time):

real time	0.04 seconds
user cpu time	0.04 seconds
system cpu time	0.01 seconds
memory	7877.90k
OS Memory	30272.00k
Timestamp	21/02/2024 09:48:09 AM
Step Count	30 Switch Count 1
Page Faults	0
Page Reclaims	2006
Page Swaps	0
Voluntary Context Switches	34
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

142
143 proc freq data=gp;
144 table sex cob healthcare_card adverse_reaction Smoke_current_GP Risky_alcohol_GP Obese_GP HighBP_GP reason/nocum;
145 run;

NOTE: There were 5300 observations read from the data set WORK.GP.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.05 seconds
user cpu time	0.05 seconds
system cpu time	0.00 seconds
memory	1948.50k
OS Memory	25128.00k
Timestamp	21/02/2024 09:48:09 AM
Step Count	31 Switch Count 2
Page Faults	0
Page Reclaims	468
Page Swaps	0
Voluntary Context Switches	12
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	280

146
147 * Univariate analysis;
148 ods graphics/ height=500px width=500px;
149 proc univariate data=gp noprint;
150 hist age drinks_day height weight syst_bp diast_bp BMI_GP /normal;
151 run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time	2.96 seconds
user cpu time	0.44 seconds
system cpu time	0.05 seconds
memory	15030.18k
OS Memory	37848.00k
Timestamp	21/02/2024 09:48:12 AM
Step Count	32 Switch Count 0
Page Faults	5
Page Reclaims	6389
Page Swaps	0
Voluntary Context Switches	5479
Involuntary Context Switches	2
Block Input Operations	1032
Block Output Operations	2552

152
153 * Bivariate Analysis??????;
154 proc sgpanel data=gp;
155 panelby sex;
156 vbar HighBP_GP /
157 response=Obese_GP
158 stat=mean;
159 run;

NOTE: PROCEDURE SG PANEL used (Total process time):

real time	0.25 seconds
user cpu time	0.07 seconds
system cpu time	0.00 seconds
memory	7391.53k
OS Memory	42176.00k
Timestamp	21/02/2024 09:48:13 AM
Step Count	33 Switch Count 17
Page Faults	8
Page Reclaims	2404
Page Swaps	0

Voluntary Context Switches	237
Involuntary Context Switches	0
Block Input Operations	944
Block Output Operations	6328

NOTE: There were 5300 observations read from the data set WORK.GP.

```
160      **** 0.2 ED data ****
161
162
163      * Check contents of dataset;
164      proc contents data=ed order=varnum; run;
```

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.03 seconds
user cpu time	0.04 seconds
system cpu time	0.01 seconds
memory	2351.62k
OS Memory	36776.00k
Timestamp	21/02/2024 09:48:13 AM
Step Count	34 Switch Count 0
Page Faults	0
Page Reclaims	291
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

```
164      !
165
166      * View the first 10 rows of data;
167      proc print data=ed (obs=10);
168      run;
```

NOTE: There were 10 observations read from the data set WORK.ED.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	2071.31k
OS Memory	36516.00k
Timestamp	21/02/2024 09:48:13 AM
Step Count	35 Switch Count 0
Page Faults	0
Page Reclaims	255
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

```
169
170      * Descriptive statistics of dataset and check missing values for numeric variables;
171      proc means data=ed mean median mode std var min max nmiss maxdec=2;
172      var ID age_ed ed_admission ed_separation;
173      run;
```

NOTE: There were 30466 observations read from the data set WORK.ED.

NOTE: PROCEDURE MEANS used (Total process time):

real time	0.03 seconds
user cpu time	0.04 seconds
system cpu time	0.00 seconds
memory	8300.06k
OS Memory	42424.00k
Timestamp	21/02/2024 09:48:13 AM
Step Count	36 Switch Count 1
Page Faults	0
Page Reclaims	1674
Page Swaps	0
Voluntary Context Switches	45
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

```
174
175      * Descriptive statistics of dataset and check missing values for categorial variables;
176      proc freq data=ed;
177      table sex_ed cob_ed interpreter health_insurance triage_category separation_mode/nocum ;
178      run;
```

NOTE: There were 30466 observations read from the data set WORK.ED.

NOTE: PROCEDURE FREQ used (Total process time):

```

real time      0.03 seconds
user cpu time  0.04 seconds
system cpu time 0.00 seconds
memory        2352.50k
OS Memory     37544.00k
Timestamp      21/02/2024 09:48:13 AM
Step Count      37  Switch Count  2
Page Faults    0
Page Reclaims   322
Page Swaps      0
Voluntary Context Switches 15
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

179
180 * Check missing values in dx1-dx5;
181 proc sql;
182 select nmiss(dx1) as dx1,nmiss(dx2) as dx2,nmiss(dx3) as dx3,nmiss(dx4) as dx4,nmiss(dx5) as dx5
183 from ed;
184 quit;
NOTE: PROCEDURE SQL used (Total process time):
real time      0.01 seconds
user cpu time  0.01 seconds
system cpu time 0.00 seconds
memory        6984.65k
OS Memory     42408.00k
Timestamp      21/02/2024 09:48:13 AM
Step Count      38  Switch Count  0
Page Faults    0
Page Reclaims   298
Page Swaps      0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 16

185
186 * Univariate analysis;
187 ods graphics on;
188 proc freq data=ed ;
189 tables sex_ed cob_ed interpreter health_insurance triage_category separation_mode/ plots=freqplot;
190 run;

NOTE: There were 30466 observations read from the data set WORK.ED.
NOTE: PROCEDURE FREQ used (Total process time):
real time      0.39 seconds
user cpu time  0.16 seconds
system cpu time 0.02 seconds
memory        3995.43k
OS Memory     38280.00k
Timestamp      21/02/2024 09:48:13 AM
Step Count      39  Switch Count  2
Page Faults    0
Page Reclaims   1484
Page Swaps      0
Voluntary Context Switches 728
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 1480

191
192 proc univariate data=ed noprint;
193 hist age_ed/normal;
194 run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
real time      0.20 seconds
user cpu time  0.11 seconds
system cpu time 0.01 seconds
memory        8342.12k
OS Memory     39256.00k
Timestamp      21/02/2024 09:48:13 AM
Step Count      40  Switch Count  0
Page Faults    0
Page Reclaims   964
Page Swaps      0
Voluntary Context Switches 810
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 408

```

```

195      ***** 0.3 PBS data *****
196
197      * Check contents of dataset;
198      proc contents data=pbs order=varnum; run;
199
NOTE: PROCEDURE CONTENTS used (Total process time):
real time      0.02 seconds
user cpu time   0.02 seconds
system cpu time 0.00 seconds
memory          2280.93k
OS Memory       37288.00k
Timestamp        21/02/2024 09:48:13 AM
Step Count        41  Switch Count  0
Page Faults      0
Page Reclaims    287
Page Swaps       0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 24

199      !
                                         * 3164 records and 6 variables are observed;

200
201      * View the first 10 rows of data;
202      proc print data=pbs (obs=10);
203      run;
204
NOTE: There were 10 observations read from the data set WORK.PBS.
NOTE: PROCEDURE PRINT used (Total process time):
real time      0.01 seconds
user cpu time   0.02 seconds
system cpu time 0.00 seconds
memory          2028.75k
OS Memory       37028.00k
Timestamp        21/02/2024 09:48:13 AM
Step Count        42  Switch Count  0
Page Faults      0
Page Reclaims    255
Page Swaps       0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

204
205      * Descriptive statistics of dataset and check missing values for numeric variables;
206      proc means data=pbs mean median mode std var min max nmiss maxdec=2;
207      var ID supply_date;
208      run;
209
NOTE: There were 3164 observations read from the data set WORK.PBS.
NOTE: PROCEDURE MEANS used (Total process time):
real time      0.02 seconds
user cpu time   0.02 seconds
system cpu time 0.00 seconds
memory          8033.93k
OS Memory       42168.00k
Timestamp        21/02/2024 09:48:13 AM
Step Count        43  Switch Count  1
Page Faults      0
Page Reclaims    1549
Page Swaps       0
Voluntary Context Switches 25
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 16

209
210      * Find unique values in character type variables;
211      proc sql;
212      select count(distinct 'ATC'\n) as 'ATC'\n,
213      count(distinct 'drug_name'\n) as 'drug_name'\n,
214      count(distinct 'item_code'\n) as 'item_code'\n,
215      count(distinct 'form_strength'\n) as 'form_strength'\n
216      from pbs;
217      quit;
NOTE: PROCEDURE SQL used (Total process time):
real time      0.01 seconds
user cpu time   0.01 seconds
system cpu time 0.01 seconds
memory          6978.21k
OS Memory       42152.00k

```

```

Timestamp          21/02/2024 09:48:13 AM
Step Count          44  Switch Count  0
Page Faults        0
Page Reclaims      327
Page Swaps         0
Voluntary Context Switches 2
Involuntary Context Switches 0
Block Input Operations 8
Block Output Operations 32

218
219 * Check missing value in character type variables;
220 proc sql;
221 select nmiss(ATC) as ATC, nmiss(drug_name) as drug_name, nmiss(item_code) as item_code,nmiss(form_strength) as
221 ! form_strength
222 from pbs;
223 quit;
NOTE: PROCEDURE SQL used (Total process time):
real time          0.00 seconds
user cpu time       0.00 seconds
system cpu time    0.00 seconds
memory             6978.96k
OS Memory          42152.00k
Timestamp          21/02/2024 09:48:13 AM
Step Count          45  Switch Count  0
Page Faults        0
Page Reclaims      259
Page Swaps         0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 8

224
225 * Univariate analysis;
226 ods graphics on;
227 proc freq data=ed ;
228 tables sex_ed cob_ed interpreter health_insurance triage_category separation_mode/ plots=freqplot;
229 run;

NOTE: There were 30466 observations read from the data set WORK.ED.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.33 seconds
user cpu time       0.18 seconds
system cpu time    0.02 seconds
memory             3933.06k
OS Memory          38024.00k
Timestamp          21/02/2024 09:48:14 AM
Step Count          46  Switch Count  2
Page Faults        0
Page Reclaims      1439
Page Swaps         0
Voluntary Context Switches 722
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 1488

230
231 proc univariate data=ed noprint;
232 hist age_ed/normal;
233 run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
real time          0.18 seconds
user cpu time       0.11 seconds
system cpu time    0.01 seconds
memory             8286.31k
OS Memory          39256.00k
Timestamp          21/02/2024 09:48:14 AM
Step Count          47  Switch Count  0
Page Faults        0
Page Reclaims      973
Page Swaps         0
Voluntary Context Switches 811
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 400

234
235
236 ****
237 ****

```

```

238     /* 1. Research Question 1 (Primary care perspective) */
239     ****;
240
241     /* 1.1. Create a new variable Agegroup_GP */
242
243     * Create a format for new variable;
244     proc format;
245
246     ! value agegroup
247     1='Under 60 years old' 2='60 years old and older';
NOTE: Format AGEGROUP has been output.
247     run;

NOTE: PROCEDURE FORMAT used (Total process time):
real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         249.09k
OS Memory       36000.00k
Timestamp       21/02/2024 09:48:14 AM
Step Count       48  Switch Count  0
Page Faults     0
Page Reclaims   14
Page Swaps      0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

248     * Create a new variable and apply format;
249     data gp1;
250     set gp;
251     * AgegroupGP =1 : Under 60 years old, AgegroupGP =2 : 60 years old and older;
252     if age < 60 then Agegroup_GP =1;
253     if age >=60 then Agegroup_GP =2;
254     format Agegroup_GP agegroup.;
255     run;

NOTE: There were 5300 observations read from the data set WORK.GP.
NOTE: The data set WORK.GP1 has 5300 observations and 19 variables.
NOTE: DATA statement used (Total process time):
real time      0.00 seconds
user cpu time   0.01 seconds
system cpu time 0.00 seconds
memory         2575.37k
OS Memory       37160.00k
Timestamp       21/02/2024 09:48:14 AM
Step Count       49  Switch Count  2
Page Faults     0
Page Reclaims   261
Page Swaps      0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 1808

256     * Check new variable created;
257     proc freq data=gp1; table Agegroup_GP /missing;
258     run;

NOTE: There were 5300 observations read from the data set WORK.GP1.
NOTE: PROCEDURE FREQ used (Total process time):
real time      0.01 seconds
user cpu time   0.01 seconds
system cpu time 0.00 seconds
memory         1611.06k
OS Memory       37160.00k
Timestamp       21/02/2024 09:48:14 AM
Step Count       50  Switch Count  2
Page Faults     0
Page Reclaims   286
Page Swaps      0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

259     /* 1.2. Calculate the proportion of GP patients who attended the ED in 2014 and interpret the finding */
260
261     * Merge gp1 and ed data and add a new variable;
262     data merge1;
263     merge gp1(in=a) ed(in=b); by ID;

```

```

265      if b then ed_record=1;
266      else ed_record=0;
267      format ed_record ynf.;
268      if a;
269      run;

NOTE: There were 5300 observations read from the data set WORK.GP1.
NOTE: There were 30466 observations read from the data set WORK.ED.
NOTE: The data set WORK.MERGE1 has 10109 observations and 34 variables.
NOTE: DATA statement used (Total process time):
      real time          0.01 seconds
      user cpu time     0.01 seconds
      system cpu time   0.01 seconds
      memory            4870.84k
      OS Memory         39980.00k
      Timestamp         21/02/2024 09:48:14 AM
      Step Count        51  Switch Count  2
      Page Faults       0
      Page Reclaims     745
      Page Swaps        0
      Voluntary Context Switches 17
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 5128

270
271      * Calculate the proportion of GP patients who attended ED in 2014;
272      proc sort data=merge1 out=merge1_nodup nodupkey; * output dataset is excluded duplicated ID;
273      by id ed_record;
274
275

NOTE: There were 10109 observations read from the data set WORK.MERGE1.
NOTE: 4809 observations with duplicate key values were deleted.
NOTE: The data set WORK.MERGE1_NODUP has 5300 observations and 34 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time          0.00 seconds
      user cpu time     0.00 seconds
      system cpu time   0.00 seconds
      memory            5913.21k
      OS Memory         40640.00k
      Timestamp         21/02/2024 09:48:14 AM
      Step Count        52  Switch Count  2
      Page Faults       0
      Page Reclaims     1084
      Page Swaps        0
      Voluntary Context Switches 15
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 2832

276      proc freq data=merge1_nodup;
277      table ed_record;
278      run;

NOTE: There were 5300 observations read from the data set WORK.MERGE1_NODUP.
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      user cpu time     0.01 seconds
      system cpu time   0.01 seconds
      memory            2175.75k
      OS Memory         37288.00k
      Timestamp         21/02/2024 09:48:14 AM
      Step Count        53  Switch Count  2
      Page Faults       0
      Page Reclaims     319
      Page Swaps        0
      Voluntary Context Switches 15
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 264

279
280      /* 1.3. Calculate total number of monthly ED admissions for all GP patients. Plot the result to show monthly trends and
281      ! interpret the results. */
282
283      * Create a new variable month that contains month of ed_admission date;
284      data merge2;
285      set merge1_nodup;
286      month = month(ed_admission);
286      run;

NOTE: Missing values were generated as a result of performing an operation on missing values.

```

Each place is given by: (Number of times) at (Line):(Column).

4017 at 285:9

NOTE: There were 5300 observations read from the data set WORK.MERGE1_NODUP.

NOTE: The data set WORK.MERGE2 has 5300 observations and 35 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	3640.78k
OS Memory	38824.00k
Timestamp	21/02/2024 09:48:14 AM
Step Count	54 Switch Count 2
Page Faults	0
Page Reclaims	532
Page Swaps	0
Voluntary Context Switches	17
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2824

287
288 * Calculate total number of monthly ED admissions for all GP patients;
289 proc sql;
290 select count(*) as Total_number from merge2 where ed_record=1;
291 select count(*) as Total_number from merge2 where ed_record=0; * no. of GP patients who don't have ED admision record
291 !
291 ! for data cross check;
292 quit;

NOTE: PROCEDURE SQL used (Total process time):

real time	0.01 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	7096.31k
OS Memory	42152.00k
Timestamp	21/02/2024 09:48:14 AM
Step Count	55 Switch Count 34
Page Faults	0
Page Reclaims	503
Page Swaps	0
Voluntary Context Switches	91
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

293
294 * Count number of ed_admission date of all GP patients by monthly;
295 proc sql;
296 create table monthly as
297 select month, count(*) as MonAdm from merge2
298 group by month;

NOTE: Table WORK.MONTHLY created, with 13 rows and 2 columns.

299 quit;

NOTE: PROCEDURE SQL used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.01 seconds
memory	6964.06k
OS Memory	42152.00k
Timestamp	21/02/2024 09:48:14 AM
Step Count	56 Switch Count 2
Page Faults	0
Page Reclaims	416
Page Swaps	0
Voluntary Context Switches	23
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	288

300
301 * Plot monthly ed_admission ;
302 title 'Distribution of monthly ED admission for all GP patients';
303 proc sgplot data=monthly;
304 vbar month / response=MonAdm ;
305 xaxis display=(nolabel)
306 values=(1 2 3 4 5 6 7 8 9 10 11 12)
307 valuesdisplay=('Jan' 'Feb' 'Mar' 'Apr' 'May' 'Jun' 'Jul' 'Aug' 'Sep' 'Oct' 'Nov' 'Dec');
308 yaxis grid label='ED admission';
309 run;

NOTE: Since no format is assigned, the numeric category variable will use the default of BEST6.

NOTE: PROCEDURE SGPLT used (Total process time):

real time	0.08 seconds
-----------	--------------

```

user cpu time      0.03 seconds
system cpu time   0.00 seconds
memory           2064.04k
OS Memory         37288.00k
Timestamp         21/02/2024 09:48:14 AM
Step Count          57  Switch Count  2
Page Faults        0
Page Reclaims      505
Page Swaps         0
Voluntary Context Switches 136
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 384

```

NOTE: There were 13 observations read from the data set WORK.MONTHLY.

```

310
311 /* 1.4. Examine differences between patients who did and did not attend the ED in 2014 */
312 /* in terms of socio-demographic characteristics [sex, age group, country of birth, and health care card] */
313 /* and health-related factors [smoking, risky alcohol consumption, obesity, and high blood pressure] */
314
315 * Compare socio-demographic characteristics and health-related factors by ed_record;
316 proc freq data=merge2;
317 tables (sex agegroup_gp cob healthcare_card smoke_current_gp risky_alcohol_gp obese_gp highbp_gp) * ed_record /chisq
318 ! norow nopercent ;
318 run;

```

NOTE: There were 5300 observations read from the data set WORK.MERGE2.

NOTE: PROCEDURE FREQ used (Total process time):

```

real time      0.16 seconds
user cpu time   0.17 seconds
system cpu time 0.01 seconds
memory           3275.40k
OS Memory         37548.00k
Timestamp         21/02/2024 09:48:14 AM
Step Count          58  Switch Count  4
Page Faults        2
Page Reclaims      525
Page Swaps         0
Voluntary Context Switches 26
Involuntary Context Switches 1
Block Input Operations 328
Block Output Operations 608

```

```

319
320 /* 1.5. Among GP patients who visited the ED, calculate the total number of ED attendance for each person in 2004. */
321 /* Describe the distribution of numbers of ED attendance using a histogram and descriptive statistics. */
322
323 * GP patients who has ED admission record;
324 data attendance; set merge1;
325 where ed_record eq 1;
326 run;

```

NOTE: There were 6092 observations read from the data set WORK.MERGE1.

WHERE ed_record=1;

NOTE: The data set WORK.ATTENDANCE has 6092 observations and 34 variables.

NOTE: DATA statement used (Total process time):

```

real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory           3516.40k
OS Memory         39336.00k
Timestamp         21/02/2024 09:48:14 AM
Step Count          59  Switch Count  52
Page Faults        0
Page Reclaims      580
Page Swaps         0
Voluntary Context Switches 147
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3080

```

```

327
328 * Count number of ED attendance per each ID;
329 proc sort data=attendance ;
330 by id; run;

```

NOTE: There were 6092 observations read from the data set WORK.ATTENDANCE.

NOTE: The data set WORK.ATTENDANCE has 6092 observations and 34 variables.

NOTE: PROCEDURE SORT used (Total process time):

```

real time      0.00 seconds
user cpu time   0.01 seconds
system cpu time 0.01 seconds
memory           4476.34k

```

```

OS Memory      39352.00k
Timestamp      21/02/2024 09:48:14 AM
Step Count      60  Switch Count  2
Page Faults    0
Page Reclaims   800
Page Swaps     0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3088

```

```

331      data att_count;
332      set attendance;
333      by id;
334      * Generate admseq - sequence within each person ID;
335      retain admseq;
336      if first.id then admseq=1;
337      else admseq = admseq + 1;
338      * Produce output with the last number of admseq of each ID;
339      if last.id then output;
340      run;
341

```

NOTE: There were 6092 observations read from the data set WORK.ATTENDANCE.

NOTE: The data set WORK.ATT_COUNT has 1283 observations and 35 variables.

NOTE: DATA statement used (Total process time):

```

real time      0.00 seconds
user cpu time  0.00 seconds
system cpu time 0.00 seconds
memory        2706.65k
OS Memory      37544.00k
Timestamp      21/02/2024 09:48:14 AM
Step Count      61  Switch Count  2
Page Faults    0
Page Reclaims   338
Page Swaps     0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 776

```

```

342      * Describe distribution of number of ED admission;
343      proc means data=att_count (keep= id admseq) min median mean max q1 q3 p10 p25 p50 p75 p90 p99;
344      run;
345

```

NOTE: There were 1283 observations read from the data set WORK.ATT_COUNT.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time      0.02 seconds
user cpu time  0.03 seconds
system cpu time 0.00 seconds
memory        7084.65k
OS Memory      41528.00k
Timestamp      21/02/2024 09:48:14 AM
Step Count      62  Switch Count  1
Page Faults    0
Page Reclaims   1404
Page Swaps     0
Voluntary Context Switches 29
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

```

```

346      * Plot the distribution;
347      title 'Distribution of number of ED admission per single GP patient';
348      proc univariate data=att_count noprint;
349      histogram admseq / normal;
350      inset N = 'Number of GP patients' Mean(8.2) Median(8.2) STD = 'Standard Deviation' (8.3) / position=ne;
351      label admseq='Number of ED adimission' ;
352      run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time      0.19 seconds
user cpu time  0.07 seconds
system cpu time 0.01 seconds
memory        7365.21k
OS Memory      38616.00k
Timestamp      21/02/2024 09:48:14 AM
Step Count      63  Switch Count  0
Page Faults    4
Page Reclaims   922
Page Swaps     0
Voluntary Context Switches 780

```

```

Involuntary Context Switches      0
Block Input Operations          584
Block Output Operations         432

353
354     /* 1.6. Continue with the results of step 1.5, select GP patients who had a large number of ED visits (i.e. top
355     ! quartile).
356     Examine and report socio-demographic and health-related characteristics of these patients */
357
358     * Select GP patients in top quartile (top 25 % of ED admission counts);
359     data top_count; set att_count;
360     where admseq > 6; /*75% quantile in ascending order is top 25% in descending order;
360     run;

NOTE: There were 318 observations read from the data set WORK.ATT_COUNT.
WHERE admseq>6;
NOTE: The data set WORK.TOP_COUNT has 318 observations and 35 variables.
NOTE: DATA statement used (Total process time):
real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory             1337.87k
OS Memory          36648.00k
Timestamp           21/02/2024 09:48:14 AM
Step Count          64  Switch Count  5
Page Faults        0
Page Reclaims      137
Page Swaps         0
Voluntary Context Switches  24
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

361
362     title 'Socio-demographic report of the highest 25% of ED visit patients';
363     proc freq data=top_count;
364     tables sex agegroup_gp cob healthcare_card /plots=freqplot ;
365     run;

NOTE: There were 318 observations read from the data set WORK.TOP_COUNT.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.20 seconds
user cpu time      0.11 seconds
system cpu time    0.01 seconds
memory             2507.18k
OS Memory          37256.00k
Timestamp           21/02/2024 09:48:15 AM
Step Count          65  Switch Count  2
Page Faults        0
Page Reclaims      886
Page Swaps         0
Voluntary Context Switches  492
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 896

366
367     title 'Health-related characteristics report of the highest 25% of ED visit patients';
368     proc freq data=top_count;
369     tables smoke_current_gp risky_alcohol_gp obese_gp highbp_gp/plots=freqplot ;
370     run;

NOTE: There were 318 observations read from the data set WORK.TOP_COUNT.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.19 seconds
user cpu time      0.10 seconds
system cpu time    0.01 seconds
memory             2506.03k
OS Memory          37256.00k
Timestamp           21/02/2024 09:48:15 AM
Step Count          66  Switch Count  2
Page Faults        0
Page Reclaims      879
Page Swaps         0
Voluntary Context Switches  487
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 896

371
372     ****
373     /* 2. Research Question 2 (ED care perspective) */

```

```

374 ****;
375
376 /* 2.1. Create three variables to flag ED records with these behaviours being recorded in any diagnosis field */
377
378 * Create a format of new variables;
379 proc format;
380
381 ! value smoker
382 0='No' 1='Yes, smoker';
NOTE: Format SMOKER has been output.
382
382 ! value drinker
383 0='No' 1='Yes, drinker';
NOTE: Format DRINKER has been output.
384
384 ! value obese
385 0='No' 1='Yes, obese';
NOTE: Format OBESE has been output.
386 run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.00 seconds
      system cpu time  0.00 seconds
      memory        272.12k
      OS Memory     36000.00k
      Timestamp     21/02/2024 09:48:15 AM
      Step Count    67  Switch Count  0
      Page Faults   0
      Page Reclaims 14
      Page Swaps    0
      Voluntary Context Switches  0
      Involuntary Context Switches 0
      Block Input Operations  0
      Block Output Operations   8

387
388 * Create new variables;
389 data ed1; set ed;
390 array diag {*} dx1-dx5;
391 smoker_flag=0;
392 risky_alcohol_flag=0;
393 obesity_flag=0;
394 do i=1 to dim(diag);
395 if diag{i} in ('F17','Z72') then smoker_flag=1;* smoker_flag = 0 : No, smoker_flag = 1 : Yes, smoker;
396 if diag{i} = 'F10' then risky_alcohol_flag=1;* risky_alcohol_flag = 0 : No, risky_alcohol_flag = 1 : Yes, drinker;
396 !
397 if diag{i} = 'E66' then obesity_flag=1;* obesity_flag = 0 : No, obesity_flag = 1 : Yes, obese;
398 drop i;
399 end;
400 format smoker_flag smoker.;
401 format risky_alcohol_flag drinker.;
402 format obesity_flag obese.;
403 run;

NOTE: There were 30466 observations read from the data set WORK.ED.
NOTE: The data set WORK.ED1 has 30466 observations and 18 variables.
NOTE: DATA statement used (Total process time):
      real time      0.02 seconds
      user cpu time  0.02 seconds
      system cpu time  0.01 seconds
      memory        3611.75k
      OS Memory     39080.00k
      Timestamp     21/02/2024 09:48:15 AM
      Step Count    68  Switch Count  2
      Page Faults   0
      Page Reclaims 548
      Page Swaps    0
      Voluntary Context Switches  11
      Involuntary Context Switches 0
      Block Input Operations  0
      Block Output Operations   7176

404
405 /* 2.2. Classify whether the patient smokes, drinks alcohol at risky level or is obese,
406 if these risk factors are recorded in any ED record for a patient.
407 Calculate and report the prevalence of smoking, risky alcohol consumption and obesity among ED patients */
408
409 * Create new variables;
410 data ed2; set ed1;
411 * Assign default value of new variables;
412 smoker_ED=0;
413 risky_alcohol_ED=0;
414 obesity_ED=0;

```

```

415      * Apply condition;
416      if smoker_flag = 1 then smoker_ED =1;
417      if risky_alcohol_flag =1 then risky_alcohol_ED =1;
418      if obesity_flag =1 then obesity_ED =1;
419      * Apply format;
420      format smoker_ED smoker.;
421      format risky_alcohol_ED drinker.;
422      format obesity_ED obese.;
423      run;

NOTE: There were 30466 observations read from the data set WORK.ED1.
NOTE: The data set WORK.ED2 has 30466 observations and 21 variables.
NOTE: DATA statement used (Total process time):
      real time      0.01 seconds
      user cpu time  0.01 seconds
      system cpu time 0.00 seconds
      memory        3596.93k
      OS Memory     39080.00k
      Timestamp     21/02/2024 09:48:15 AM
      Step Count    69  Switch Count  2
      Page Faults   0
      Page Reclaims 536
      Page Swaps    0
      Voluntary Context Switches 19
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 8720

424
425      * Summarize dataset per each ED patient ID;
426      proc sort data=ed2; by id; run;

NOTE: There were 30466 observations read from the data set WORK.ED2.
NOTE: The data set WORK.ED2 has 30466 observations and 21 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.01 seconds
      user cpu time  0.00 seconds
      system cpu time 0.01 seconds
      memory        7832.81k
      OS Memory     42696.00k
      Timestamp     21/02/2024 09:48:15 AM
      Step Count    70  Switch Count  2
      Page Faults   0
      Page Reclaims 1600
      Page Swaps    0
      Voluntary Context Switches 13
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 8712

427      proc means noprint data=ed2; by id; output out=ed3 (drop = _type_ _freq_)
428      max(smoker_ED) =smoker_ED
429      max(risky_alcohol_ED) = risky_alcohol_ED
430      max(obesity_ED) = obesity_ED;
431      run;

NOTE: There were 30466 observations read from the data set WORK.ED2.
NOTE: The data set WORK.ED3 has 5637 observations and 4 variables.
NOTE: PROCEDURE MEANS used (Total process time):
      real time      0.05 seconds
      user cpu time  0.05 seconds
      system cpu time 0.03 seconds
      memory        3378.71k
      OS Memory     38316.00k
      Timestamp     21/02/2024 09:48:15 AM
      Step Count    71  Switch Count  954
      Page Faults   0
      Page Reclaims 431
      Page Swaps    0
      Voluntary Context Switches 2393
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 520

432
433      * Calculate prevalence of smoking, risky alcohol consumption and obesity;
434      proc freq data=ed3;
435      tables smoker_ED risky_alcohol_ED obesity_ED ;
436      run;

NOTE: There were 5637 observations read from the data set WORK.ED3.
NOTE: PROCEDURE FREQ used (Total process time):
      real time      0.02 seconds

```

```
user cpu time      0.02 seconds
system cpu time   0.00 seconds
memory           1069.53k
OS Memory        36520.00k
Timestamp         21/02/2024 09:48:15 AM
Step Count        72  Switch Count  2
Page Faults      0
Page Reclaims    126
Page Swaps       0
Voluntary Context Switches 10
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264
```

```
437
438
439
440 /* 2.3. Calculate and report the proportion of ED patients who had a visit to Medical Plus GP in 2014 */
441 * Merge ed2 and gp data and add a new variable gp_record;
442 data merge3;
443 merge ed2(in=a) gp(in=b); by id;
444 if b then gp_record=1;
445 else gp_record=0;
446 format gp_record ynf. ;
447 if a;
448 run;
```

NOTE: There were 30466 observations read from the data set WORK.ED2.
NOTE: There were 5300 observations read from the data set WORK.GP.

NOTE: The data set WORK.MERGE3 has 30466 observations and 39 variables.

NOTE: DATA statement used (Total process time):

```
real time        0.01 seconds
user cpu time   0.01 seconds
system cpu time 0.01 seconds
memory          4894.21k
OS Memory       40236.00k
Timestamp        21/02/2024 09:48:15 AM
Step Count       73  Switch Count  2
Page Faults     0
Page Reclaims   770
Page Swaps      0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 17672
```

```
449
450 * Calculate proportion of ED patients who has gp_record or not;
451 proc sort data=merge3 out=merge3_nodup nodupkey;
452 by id gp_record;
453
```

NOTE: There were 30466 observations read from the data set WORK.MERGE3.

NOTE: 24829 observations with duplicate key values were deleted.

NOTE: The data set WORK.MERGE3_NODUP has 5637 observations and 39 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time        0.01 seconds
user cpu time   0.00 seconds
system cpu time 0.01 seconds
memory          13205.12k
OS Memory       48100.00k
Timestamp        21/02/2024 09:48:15 AM
Step Count       74  Switch Count  3
Page Faults     0
Page Reclaims   2796
Page Swaps      0
Voluntary Context Switches 20
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3344
```

```
454 proc freq data=merge3_nodup;
455   table gp_record;
456 run;
```

NOTE: There were 5637 observations read from the data set WORK.MERGE3_NODUP.

NOTE: PROCEDURE FREQ used (Total process time):

```
real time        0.01 seconds
user cpu time   0.01 seconds
system cpu time 0.00 seconds
memory          2197.25k
OS Memory       37288.00k
Timestamp        21/02/2024 09:48:15 AM
```

```

Step Count          75  Switch Count  2
Page Faults        0
Page Reclaims      316
Page Swaps         0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

457
458 /* 2.4. Examine whether there are any differences between ED patients who did and did not visit a GP */
459 /* in terms of sex, age, country of birth, private health insurance, smoking, risky alcohol consumption and obesity. */
460 /* You can categorise patient age into two groups (under 60 /60 and older). */
461
462 * Create a new variable agegroup_ed and apply format;
463 data merge4;
464 set merge3_nodup;
465 * AgegroupGP =1 : Under 60 years old, AgegroupGP =2 : 60 years old and older;
466 if 0< age_ed < 60 then agegroup_ed =1;
467 if age_ed =>60 then agegroup_ed =2;
468 format agegroup_ed agegroup.;
469 run;

NOTE: There were 5637 observations read from the data set WORK.MERGE3_NODUP.
NOTE: The data set WORK.MERGE4 has 5637 observations and 40 variables.
NOTE: DATA statement used (Total process time):
real time          0.00 seconds
user cpu time       0.00 seconds
system cpu time     0.00 seconds
memory              3666.50k
OS Memory           39080.00k
Timestamp            21/02/2024 09:48:15 AM
Step Count          76  Switch Count  2
Page Faults         0
Page Reclaims       536
Page Swaps          0
Voluntary Context Switches 17
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3600

470
471 proc freq data=merge4;
472   tables (sex_ed agegroup_ed cob_ed health_insurance smoker_ed risky_alcohol_ed obesity_ed) * gp_record /chisq norow
473 ! nopercent ;
473 run;

NOTE: There were 5637 observations read from the data set WORK.MERGE4.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.15 seconds
user cpu time       0.15 seconds
system cpu time     0.00 seconds
memory              2789.71k
OS Memory           37548.00k
Timestamp            21/02/2024 09:48:15 AM
Step Count          77  Switch Count  4
Page Faults         0
Page Reclaims       396
Page Swaps          0
Voluntary Context Switches 26
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 608

474
475 * Plot data;
476 proc sgpanel data=merge4;
477 panelby gp_record;
478 vbar agegroup_ed ;
479 run;

NOTE: PROCEDURE SGANEL used (Total process time):
real time          0.20 seconds
user cpu time       0.08 seconds
system cpu time     0.02 seconds
memory              9481.93k
OS Memory           45760.00k
Timestamp            21/02/2024 09:48:15 AM
Step Count          78  Switch Count  16
Page Faults         0
Page Reclaims       3070
Page Swaps          0
Voluntary Context Switches 238

```

```
Involuntary Context Switches      0
Block Input Operations          0
Block Output Operations        12088
```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

```
480
481      proc sgpanel data=merge4;
482      panelby gp_record;
483      vbar cob_ed ;
484      run;
```

NOTE: PROCEDURE SG PANEL used (Total process time):

```
real time      0.17 seconds
user cpu time   0.09 seconds
system cpu time  0.02 seconds
memory         9446.31k
OS Memory       45760.00k
Timestamp       21/02/2024 09:48:16 AM
Step Count       79  Switch Count  16
Page Faults     0
Page Reclaims   3049
Page Swaps       0
Voluntary Context Switches  224
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 12072
```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

```
485
486      proc sgpanel data=merge4;
487      panelby gp_record;
488      vbar smoker_ed ;
489      run;
```

NOTE: PROCEDURE SG PANEL used (Total process time):

```
real time      0.16 seconds
user cpu time   0.09 seconds
system cpu time  0.01 seconds
memory         9420.62k
OS Memory       45760.00k
Timestamp       21/02/2024 09:48:16 AM
Step Count       80  Switch Count  16
Page Faults     0
Page Reclaims   3083
Page Swaps       0
Voluntary Context Switches  224
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 12072
```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

```
490
491      proc sgpanel data=merge4;
492      panelby gp_record;
493      vbar risky_alcohol_ed ;
494      run;
```

NOTE: PROCEDURE SG PANEL used (Total process time):

```
real time      0.17 seconds
user cpu time   0.09 seconds
system cpu time  0.02 seconds
memory         9414.96k
OS Memory       45504.00k
Timestamp       21/02/2024 09:48:16 AM
Step Count       81  Switch Count  16
Page Faults     0
Page Reclaims   2974
Page Swaps       0
Voluntary Context Switches  221
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 12072
```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

```
495
496      proc sgpanel data=merge4;
497      panelby gp_record;
498      vbar obesity_ed ;
499      run;
```

NOTE: PROCEDURE SG PANEL used (Total process time):

```
real time      0.19 seconds
```

```

user cpu time      0.09 seconds
system cpu time   0.03 seconds
memory           9414.53k
OS Memory        45504.00k
Timestamp         21/02/2024 09:48:16 AM
Step Count          82  Switch Count  16
Page Faults        0
Page Reclaims      3004
Page Swaps         0
Voluntary Context Switches 226
Involuntary Context Switches 2
Block Input Operations 0
Block Output Operations 12072

```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

```

500
501
502 /* 2.5. Calculate overall sensitivity (Sn) and specificity (Sp) of the recording of patient smoking in the ED data,  */
503 /* using patient smoking information in the GP data as the gold standard.  */
504 /* Comment on overall quality of ED data on patient smoking */
505
506 * Create new variables - TP(true positive),TN(true negative),FP(false positive) and FN(false negative);
507 data sn_sp; set merge4;
508 if smoke_current_gp =1 then do;
509 if smoker_ed =1 then result ="TP";
510 else if smoker_ed =0 then result = "FN";
511 end;
512 else if smoke_current_gp =0 then do;
513 if smoker_ed = 1 then result = "FP";
514 else if smoker_ed = 0 then result = "TN";
515 end;
516 run;

```

NOTE: There were 5637 observations read from the data set WORK.MERGE4.

NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.

NOTE: DATA statement used (Total process time):

```

real time      0.00 seconds
user cpu time  0.01 seconds
system cpu time 0.00 seconds
memory         3682.28k
OS Memory     39080.00k
Timestamp      21/02/2024 09:48:16 AM
Step Count      83  Switch Count  2
Page Faults    0
Page Reclaims  536
Page Swaps     0
Voluntary Context Switches 13
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3592

```

```

517
518 * Count number of TP,FN,FP,TN;
519 proc sort data=sn_sp;
520 by smoke_current_gp smoker_ed;
521 run;

```

NOTE: There were 5637 observations read from the data set WORK.SN_SP.

NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.

NOTE: PROCEDURE SORT used (Total process time):

```

real time      0.00 seconds
user cpu time  0.00 seconds
system cpu time 0.00 seconds
memory         4478.21k
OS Memory     39352.00k
Timestamp      21/02/2024 09:48:16 AM
Step Count      84  Switch Count  2
Page Faults    0
Page Reclaims  821
Page Swaps     0
Voluntary Context Switches 12
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3592

```

```

522
523 data sn_sp2 (keep= tp fn fp tn);
524 set sn_sp;
525 by smoke_current_gp;
526 retain tp fn fp tn;
527 if first.smoke_current_gp then do;
528 tp=0; fn=0; fp=0; tn=0;
529 end;

```

```

530      if result in ("TP") then tp=tp+1;
531      if result in ("FN") then fn=fn+1;
532      if result in ("FP") then fp=fp+1;
533      if result in ("TN") then tn=tn+1;
534      else;
535      if last.smoke_current_gp then output;
536      run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: The data set WORK.SN_SP2 has 3 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time          0.00 seconds
      user cpu time     0.01 seconds
      system cpu time   0.00 seconds
      memory            2439.71k
      OS Memory         37288.00k
      Timestamp         21/02/2024 09:48:16 AM
      Step Count        85   Switch Count  2
      Page Faults       0
      Page Reclaims     323
      Page Swaps        0
      Voluntary Context Switches 11
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 264

537      * Create a table of counted elements;
538      proc sql;
539      create table sn_sp3 as
540      select sum(tp) as TP, sum(fn) as FN, sum(fp) as FP, sum(tn) as TN
541      from sn_sp2;
542
NOTE: Table WORK.SN_SP3 created, with 1 rows and 4 columns.

543      quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time          0.00 seconds
      user cpu time     0.00 seconds
      system cpu time   0.00 seconds
      memory            5648.84k
      OS Memory         41384.00k
      Timestamp         21/02/2024 09:48:16 AM
      Step Count        86   Switch Count  2
      Page Faults       0
      Page Reclaims     123
      Page Swaps        0
      Voluntary Context Switches 12
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 264

544      * Calculate sensitivity and specificity;
545      proc sql;
546      create table sn_sp4 as
547      select tp/(tp+fn) as Sensitivity, tn/(tn+fp) as Specificity, (tn+tp)/(tn+tp+fn+fp) as Accuracy
548      from sn_sp3;
549
NOTE: Table WORK.SN_SP4 created, with 1 rows and 3 columns.

550      quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time          0.00 seconds
      user cpu time     0.00 seconds
      system cpu time   0.00 seconds
      memory            5620.43k
      OS Memory         41384.00k
      Timestamp         21/02/2024 09:48:16 AM
      Step Count        87   Switch Count  2
      Page Faults       0
      Page Reclaims     125
      Page Swaps        0
      Voluntary Context Switches 13
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 264

551
552
553      /* 2.6. Repeat calculation of Sn and Sp of the recording of smoking in ED data, */
554      /* separately for each patient's sex, age group, country of birth and private health insurance (i.e. stratified by */
555      ! sociodemographic factors). */
556      /* Comment on whether recording of smoking information in ED data differs by patient sociodemographic characteristics */

```

```

557      * Calculate Sn, Sp by sex;
558      proc sort data=sn_sp; by sex_ed; run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.01 seconds
      system cpu time 0.01 seconds
      memory        4481.93k
      OS Memory     39352.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count     88   Switch Count  2
      Page Faults    0
      Page Reclaims  809
      Page Swaps     0
      Voluntary Context Switches 16
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 3592

559
560      proc freq data=sn_sp;
561      by sex_ed;
562      tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
563      run;

NOTE: No statistics are computed for smoker_ED * Smoke_current_GP because smoker_ED has fewer than 2 nonmissing levels.
NOTE: The above message was for the following BY group:
      Sex of patient=.
NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: PROCEDURE FREQ used (Total process time):
      real time      0.05 seconds
      user cpu time  0.05 seconds
      system cpu time 0.00 seconds
      memory        3061.03k
      OS Memory     37548.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count     89   Switch Count  4
      Page Faults    4
      Page Reclaims  407
      Page Swaps     0
      Voluntary Context Switches 33
      Involuntary Context Switches 0
      Block Input Operations 576
      Block Output Operations 544

564      * Calculate Sn, Sp by age group;
565      proc sort data=sn_sp; by agegroup_ed; run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.00 seconds
      system cpu time 0.00 seconds
      memory        4484.59k
      OS Memory     39352.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count     90   Switch Count  2
      Page Faults    0
      Page Reclaims  810
      Page Swaps     0
      Voluntary Context Switches 11
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 3592

567
568      proc freq data=sn_sp;
569      by agegroup_ed;
570      tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
571      run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: PROCEDURE FREQ used (Total process time):
      real time      0.04 seconds
      user cpu time  0.04 seconds
      system cpu time 0.01 seconds
      memory        2591.34k
      OS Memory     37548.00k
      Timestamp     21/02/2024 09:48:16 AM

```

```

Step Count          91  Switch Count  4
Page Faults        0
Page Reclaims      387
Page Swaps         0
Voluntary Context Switches 23
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 560

572
573     * Calculate Sn, Sp by country of birth;
574     proc sort data=sn_sp; by cob_ed; run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.00 seconds
      system cpu time 0.00 seconds
      memory        4484.12k
      OS Memory     39352.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count    92  Switch Count  2
      Page Faults   0
      Page Reclaims 811
      Page Swaps    0
      Voluntary Context Switches 12
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 3592

575
576     proc freq data=sn_sp;
577     by cob_ed;
578     tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
579     run;

NOTE: No statistics are computed for smoker_ED * Smoke_current_GP because all data are missing.
NOTE: The above message was for the following BY group:
      Country of birth=.
NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: PROCEDURE FREQ used (Total process time):
      real time      0.04 seconds
      user cpu time  0.04 seconds
      system cpu time 0.01 seconds
      memory        2765.84k
      OS Memory     37548.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count    93  Switch Count  4
      Page Faults   0
      Page Reclaims 391
      Page Swaps    0
      Voluntary Context Switches 23
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 544

580
581     * Calculate Sn, Sp by private health insurance;
582     proc sort data=sn_sp; by health_insurance; run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: The data set WORK.SN_SP has 5637 observations and 41 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time      0.00 seconds
      user cpu time  0.00 seconds
      system cpu time 0.00 seconds
      memory        4480.00k
      OS Memory     39352.00k
      Timestamp     21/02/2024 09:48:16 AM
      Step Count    94  Switch Count  2
      Page Faults   0
      Page Reclaims 810
      Page Swaps    0
      Voluntary Context Switches 12
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 3592

583
584     proc freq data=sn_sp;
585     by health_insurance;

```

```

586      tables smoker_ed * smoke_current_gp/nopercent nocol norow senspec;
587      run;

NOTE: There were 5637 observations read from the data set WORK.SN_SP.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.06 seconds
user cpu time       0.06 seconds
system cpu time    0.01 seconds
memory             2588.96k
OS Memory          37548.00k
Timestamp          21/02/2024 09:48:16 AM
Step Count          95  Switch Count  4
Page Faults         0
Page Reclaims       386
Page Swaps          0
Voluntary Context Switches 28
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 560

588
589 ** Note: in above calculation using 'senspec' reference cell is set TN instead of TP,
590 therefore sensitivity and specificity value in result table are displayed in opposite way.
591 I tried to apply table option 'SENSPEC REFCELL=' but could not produce the result by getting syntax error;
592 ****
593 /* 3. Research Question 3 (Tobacco control perspective) */
594 ****;
595 ****;
596
597 /* 3.1. Create a cohort of Sunnydale residents who smoke using information from the GP and ED data sources. */
598 /* How many smokers could you identify in the GP data alone, ED data alone, and a combination of both GP/ED data sources
599 ! */
600
601 * Identify unique ID without duplicated in GP and ED data;
proc sort data=gp1 out=gp_id nodupkey; by id; run;

NOTE: There were 5300 observations read from the data set WORK.GP1.
NOTE: 0 observations with duplicate key values were deleted.
NOTE: The data set WORK.GP_ID has 5300 observations and 19 variables.
NOTE: PROCEDURE SORT used (Total process time):
real time          0.00 seconds
user cpu time       0.00 seconds
system cpu time    0.00 seconds
memory             4182.43k
OS Memory          39484.00k
Timestamp          21/02/2024 09:48:16 AM
Step Count          96  Switch Count  2
Page Faults         0
Page Reclaims       482
Page Swaps          0
Voluntary Context Switches 11
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 1808

602 proc sort data=ed2 out=ed_id nodupkey; by id; run;

NOTE: There were 30466 observations read from the data set WORK.ED2.
NOTE: 24829 observations with duplicate key values were deleted.
NOTE: The data set WORK.ED_ID has 5637 observations and 21 variables.
NOTE: PROCEDURE SORT used (Total process time):
real time          0.01 seconds
user cpu time       0.02 seconds
system cpu time    0.00 seconds
memory             8103.00k
OS Memory          42956.00k
Timestamp          21/02/2024 09:48:16 AM
Step Count          97  Switch Count  2
Page Faults         0
Page Reclaims       1508
Page Swaps          0
Voluntary Context Switches 14
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 1808

603
604 * Merge GP and Ed;
605 data merge5 ;
606 merge gp_id(in=a) ed_id(in=b);
607 if a or b;
608 by id;
609 run;

```

NOTE: There were 5300 observations read from the data set WORK.GP_ID.
 NOTE: There were 5637 observations read from the data set WORK.ED_ID.
 NOTE: The data set WORK.MERGE5 has 9654 observations and 39 variables.
 NOTE: DATA statement used (Total process time):
 real time 0.00 seconds
 user cpu time 0.00 seconds
 system cpu time 0.01 seconds
 memory 4345.12k
 OS Memory 39340.00k
 Timestamp 21/02/2024 09:48:16 AM
 Step Count 98 Switch Count 2
 Page Faults 0
 Page Reclaims 636
 Page Swaps 0
 Voluntary Context Switches 12
 Involuntary Context Switches 0
 Block Input Operations 0
 Block Output Operations 5640

610 * Check if there are duplicated id;
 611 proc sort data=merge5 out=dup_check nodupkey; by id; run;
 NOTE: There were 9654 observations read from the data set WORK.MERGE5.
 NOTE: 0 observations with duplicate key values were deleted.
 NOTE: The data set WORK.DUP_CHECK has 9654 observations and 39 variables.
 NOTE: PROCEDURE SORT used (Total process time):
 real time 0.00 seconds
 user cpu time 0.01 seconds
 system cpu time 0.00 seconds
 memory 5908.90k
 OS Memory 40640.00k
 Timestamp 21/02/2024 09:48:16 AM
 Step Count 99 Switch Count 2
 Page Faults 0
 Page Reclaims 1130
 Page Swaps 0
 Voluntary Context Switches 15
 Involuntary Context Switches 0
 Block Input Operations 0
 Block Output Operations 5648

613 * Count number of smokers identified in each dataset;
 614 proc sql;
 615 select count(id) as gp_smoker from gp_id where smoke_current_gp eq 1;
 616 select count(id) as ed_smoker from ed_id where smoker_ed eq 1;
 617 select count(id) as gped_smoker from merge5 where (smoke_current_gp eq 1) or (smoker_ed eq 1);
 618 quit;
 NOTE: PROCEDURE SQL used (Total process time):
 real time 0.01 seconds
 user cpu time 0.02 seconds
 system cpu time 0.01 seconds
 memory 7183.50k
 OS Memory 42152.00k
 Timestamp 21/02/2024 09:48:16 AM
 Step Count 100 Switch Count 18
 Page Faults 0
 Page Reclaims 663
 Page Swaps 0
 Voluntary Context Switches 57
 Involuntary Context Switches 0
 Block Input Operations 0
 Block Output Operations 0

620 * Cleanup merged dataset and create cohort group;
 621 proc sort data=merge5(drop= sex age age_ed cob dx1 dx2 dx3 dx4 dx5 smoker_flag risky_alcohol_flag obesity_flag)
 622 ! out=cohort nodupkey;
 623 by id;
 624 where (smoke_current_gp eq 1) or (smoker_ed eq 1); * cohort group include only smokers;
 625 run;

NOTE: There were 922 observations read from the data set WORK.MERGE5.
 WHERE (smoke_current_gp=1) or (smoker_ed=1);
 NOTE: 0 observations with duplicate key values were deleted.
 NOTE: The data set WORK.COHORT has 922 observations and 27 variables.
 NOTE: PROCEDURE SORT used (Total process time):
 real time 0.00 seconds
 user cpu time 0.00 seconds
 system cpu time 0.00 seconds
 memory 3090.96k

```

OS Memory      38068.00k
Timestamp      21/02/2024 09:48:16 AM
Step Count      101  Switch Count  11
Page Faults    0
Page Reclaims   376
Page Swaps     0
Voluntary Context Switches 41
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 536

626
627      * remove complete duplicates;
628      * keep sex_ed and drop sex (sex matches for all ID but sex_ed has more data compared to sex);
629      * keep cob_ed and drop cob
630      * drop age and age_ed and keep agegroup_gp (ther is a slight different in age and age_ed of some ID, so just using
630 ! agegroup)
631      * drop dx1-dx5
632      * drop smoker_flag risky_alcohol_flag obesity_flag
633
634      * Check cohort dataset;
635      proc contents data=cohort varnum; run;

NOTE: PROCEDURE CONTENTS used (Total process time):
real time      0.04 seconds
user cpu time   0.04 seconds
system cpu time 0.00 seconds
memory         1181.25k
OS Memory       36520.00k
Timestamp      21/02/2024 09:48:17 AM
Step Count      102  Switch Count  0
Page Faults    1
Page Reclaims   103
Page Swaps     0
Voluntary Context Switches 1
Involuntary Context Switches 0
Block Input Operations 48
Block Output Operations 40

636
637      * See the first 10 rows of the cohort data;
638      proc print data=cohort (obs=10); run;

NOTE: There were 10 observations read from the data set WORK.COHORT.
NOTE: PROCEDURE PRINT used (Total process time):
real time      0.03 seconds
user cpu time   0.03 seconds
system cpu time 0.00 seconds
memory         972.21k
OS Memory       36260.00k
Timestamp      21/02/2024 09:48:17 AM
Step Count      103  Switch Count  0
Page Faults    0
Page Reclaims   65
Page Swaps     0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 16

639
640      /* 3.2. Examine PBS data against the cohort defined in Step 3.1 and comment on the value of PBS data as an additional
640 ! data source (i.e. on top of GP and ED data) */
641      /* to identify people who smoke and who were not identified in GP or ED data */
642
643      * Classify ID in PBS data who has PBS record of N07BA01, N06AX12 or N07BA03;
644      data pbs2(keep=id atc); set pbs;
645      where ATC = 'N07BA01' or ATC = 'N06AX12' or ATC = 'N07BA03';
646      run;

NOTE: There were 315 observations read from the data set WORK.PBS.
WHERE ATC in ('N06AX12', 'N07BA01', 'N07BA03');
NOTE: The data set WORK.PBS2 has 315 observations and 2 variables.
NOTE: DATA statement used (Total process time):
real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         2330.71k
OS Memory       37288.00k
Timestamp      21/02/2024 09:48:17 AM
Step Count      104  Switch Count  5
Page Faults    0
Page Reclaims   323

```

```

Page Swaps          0
Voluntary Context Switches 24
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

647      proc freq data=pbs2;
648          tables id atc /missing;
649          run;

NOTE: There were 315 observations read from the data set WORK.PBS2.
NOTE: PROCEDURE FREQ used (Total process time):
real time      0.11 seconds
user cpu time   0.12 seconds
system cpu time 0.01 seconds
memory         875.78k
OS Memory       36520.00k
Timestamp        21/02/2024 09:48:17 AM
Step Count        105  Switch Count  2
Page Faults      0
Page Reclaims    128
Page Swaps        0
Voluntary Context Switches 11
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 344

651      * Find IDs in PBS data that is not in cohort(GP&ED) data;
652      proc sql;
653          create table missing_smoker as
654              select id
655                  from pbs2 cohort
656                  where pbs2.id not in (select id from cohort);
657      NOTE: Table WORK.MISSING_SMOKER created, with 81 rows and 1 columns.

658      quit;
NOTE: PROCEDURE SQL used (Total process time):
real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         6038.37k
OS Memory       41644.00k
Timestamp        21/02/2024 09:48:17 AM
Step Count        106  Switch Count  2
Page Faults      0
Page Reclaims    172
Page Swaps        0
Voluntary Context Switches 11
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 272

659
660
661      /* 3.3. For the cohort created in Step 3.1, calculate the proportion of smokers who used any of the smoking cessation
662      ! therapies, */
663      /* as well as each of the three individual medicine in 2014. */
664
665      * Merge cohort and PBS data;
666      data merge6;
667      merge cohort(in=a) pbs2(in=b);
668      if a;
669      by id;
run;

NOTE: There were 922 observations read from the data set WORK.COHORT.
NOTE: There were 315 observations read from the data set WORK.PBS2.
NOTE: The data set WORK.MERGE6 has 1004 observations and 28 variables.
NOTE: DATA statement used (Total process time):
real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         1733.25k
OS Memory       36780.00k
Timestamp        21/02/2024 09:48:17 AM
Step Count        107  Switch Count  2
Page Faults      0
Page Reclaims    142
Page Swaps        0
Voluntary Context Switches 14
Involuntary Context Switches 0

```

```

Block Input Operations          0
Block Output Operations        520

670
671      * Calculate proportion who used any of the smoking cessation therapies (where ATC is NULL/ATC is NOT NULL);
672      proc freq data=merge6;
673      table atc /plots=freqplot;
674      run;

NOTE: There were 1004 observations read from the data set WORK.MERGE6.
NOTE: PROCEDURE FREQ used (Total process time):
real time          0.07 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory             2550.09k
OS Memory          37256.00k
Timestamp          21/02/2024 09:48:17 AM
Step Count          108  Switch Count  2
Page Faults        0
Page Reclaims      352
Page Swaps         0
Voluntary Context Switches 158
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 616

675
676      * Calculate proportions of smokers who has ATC variable value of N07BA01, N06AX12 or N07BA03;
677      proc sql;
678      create table out as
679      select atc, (count(atc)/(select count(atc) from merge6)) *100 as percent
680      from merge6
681      group by atc;
NOTE: Table WORK.OUT created, with 3 rows and 2 columns.

682      quit;
NOTE: PROCEDURE SQL used (Total process time):
real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory             5815.09k
OS Memory          41384.00k
Timestamp          21/02/2024 09:48:17 AM
Step Count          109  Switch Count  2
Page Faults        0
Page Reclaims      163
Page Swaps         0
Voluntary Context Switches 13
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 280

683      **** End ****
684
685
686
687
688
689
690
691
692
693
694
695
696
697      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
698      ODS HTML CLOSE;
699      &GRAPHTERM; ;*';*";*/;RUN;QUIT;
700      QUIT;RUN;
701      ODS HTML5 (ID=WEB) CLOSE;
702
703      FILENAME _GSFNAME;
NOTE: Fileref _GSFNAME has been deassigned.
704      DATA _NULL_;
705      RUN;

NOTE: DATA statement used (Total process time):
real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory             468.90k
OS Memory          26536.00k

```

Timestamp 21/02/2024 09:48:17 AM
 Step Count 110 Switch Count 0
 Page Faults 0
 Page Reclaims 27
 Page Swaps 0
 Voluntary Context Switches 0
 Involuntary Context Switches 0
 Block Input Operations 0
 Block Output Operations 0

706 OPTIONS NOTES STIMER SOURCE SYNTAXCHECK;
 707

Results: code.sas

The CONTENTS Procedure

Data Set Name	WORK.GP	Observations	5300
Member Type	DATA	Variables	18
Engine	V9	Indexes	0
Created	21/02/2024 20:48:10	Observation Length	152
Last Modified	21/02/2024 20:48:10	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	7
First Data Page	1
Max Obs per Page	861
Obs in First Data Page	831
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A680001DAF4_odaws02-apse1-2.oda,sas.com/SAS_workC01E0001DAF4_odaws02-apse1-2.oda,sas.com/gp,sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	525832
Access Permission	rw-r--r--
Owner Name	u62298616
File Size	1MB
File Size (bytes)	1048576

Variables in Creation Order					
#	Variable	Type	Len	Format	Label
1	ID	Num	8		Unique patientID
2	GP_last	Num	8	DDMMYY10.	Date of most recent GP visit
3	age	Num	8		Age at the most recent GP visit
4	sex	Num	8	SEXF.	Sex of patient
5	cob	Num	8	COBF.	Country of birth
6	healthcare_card	Num	8	YNF.	Have a health care card
7	drinks_day	Num	8		Number of alcohol drinks per day
8	height	Num	8		Body height (m)
9	weight	Num	8		Body weight (kg)
10	adverse_reaction	Num	8	YNF.	Had any reaction to aby medication
11	syst_bp	Num	8		Systolic blood pressure (mmHg)
12	diast_bp	Num	8		Diastolic blood pressure (mmHg)
13	reason	Char	16		Reason for the most recent GP visit
14	Smoke_current_GP	Num	8	YNF.	Being a current smoker
15	Risky_alcohol_GP	Num	8	YNF.	Have two or more alcohol drinks per day
16	BMI_GP	Num	8	YNF.	BMI score
17	Obese_GP	Num	8	YNF.	Being obese (BMI>=30)
18	HighBP_GP	Num	8	YNF.	Have high blood pressure (>=135/85mmHg)

Obs	ID	GP_last	age	sex	cob	healthcare_card	drinks_day	height	weight	adverse_reaction	syst_bp	diast_bp	reason	Smoke_current_GP	Risky_alcohol_GP	BMI_GP	Obese_GP
1	1	27/02/2014	59	Female	Overseas	Yes	2	1.73	61.8	No	131	89	HEADACHE	No	No	21	No
2	2	24/07/2014	20	Male	Overseas	No	1	1.83	100.2	No	122	75	HEADACHE	No	No	30	No
3	3	11/04/2014	65	Male	Australia	No	0	1.86	65.0	No	132	88	ITCHING	No	No	19	No
4	4	09/04/2014	33	Male	Overseas	Yes	1	1.75	74.9	No	121	77	HEADACHE	No	No	24	No
5	5	27/02/2014	57	Female	Australia	No	0	1.69	71.8	No	116	74	HEADACHE	No	No	25	No
6	7	01/09/2014	52	Male	Australia	No	0	1.51	107.5	No	115	66	VOMITING	No	No	47	Yes
7	8	26/07/2014	55	Female	Overseas	No	0	1.72	53.1	No	115	68	HEADACHE	No	No	18	No
8	9	23/10/2014	68	Female	Australia	No	0	1.71	84.8	No	126	76	SKIN RASH	No	No	29	No
9	10	29/07/2014	41	Male	Australia	No	0	1.88	94.1	No	129	82	HEADACHE	No	No	27	No

Obs	ID	GP_last	age	sex	cob	healthcare_card	drinks_day	height	weight	adverse_reaction	syst_bp	diast_bp	reason	Smoke_current_GP	Risky_alcohol_GP	BMI_GP	Obese_GP	t
10	11	19/06/2014	32	Female	Overseas	No	1	1.74	66.8	No	102	56	HEADACHE	No	No	22	No	

The MEANS Procedure

Variable	Label	Mean	Median	Mode	Std Dev	Variance	Minimum	Maximum	N Miss
ID	Unique patientID	2901.06	2902.50	1671.62	2794318.84	1.00	5794.00	0	
GP_last	Date of most recent GP visit	19872.20	19870.50	19780.00	85.55	7318.27	19724.00	20023.00	0
age	Age at the most recent GP visit	46.96	46.00	46.00	14.14	199.98	18.00	98.00	0
drinks_day	Number of alcohol drinks per day	1.17	1.00	0.00	1.55	2.39	0.00	18.00	294
height	Body height (m)	1.72	1.73	1.73	0.08	0.01	1.45	1.89	60
weight	Body weight (kg)	78.84	75.20	104.10	20.26	410.67	44.60	140.50	44
syst_bp	Systolic blood pressure (mmHg)	120.89	120.00	140.00	15.21	231.46	88.00	181.00	0
diast_bp	Diastolic blood pressure (mmHg)	77.63	78.00	92.00	13.30	177.00	51.00	132.00	0
BMI_GP	BMI score	26.90	25.21	33.00	7.52	56.61	14.27	66.25	91

The FREQ Procedure

Sex of patient		
sex	Frequency	Percent
Male	2364	44.69
Female	2926	55.31
Frequency Missing = 10		

Country of birth		
cob	Frequency	Percent
Australia	2233	42.64
Overseas	3004	57.36
Frequency Missing = 63		

Have a health care card		
healthcare_card	Frequency	Percent
No	3604	68.10
Yes	1688	31.90
Frequency Missing = 8		

Had any reaction to aby medication		
adverse_reaction	Frequency	Percent
No	5097	96.17
Yes	203	3.83

Being a current smoker		
Smoke_current_GP	Frequency	Percent
No	4495	85.59
Yes	757	14.41
Frequency Missing = 48		

Have two or more alcohol drinks per day		
Risky_alcohol_GP	Frequency	Percent
No	4083	81.56
Yes	923	18.44
Frequency Missing = 294		

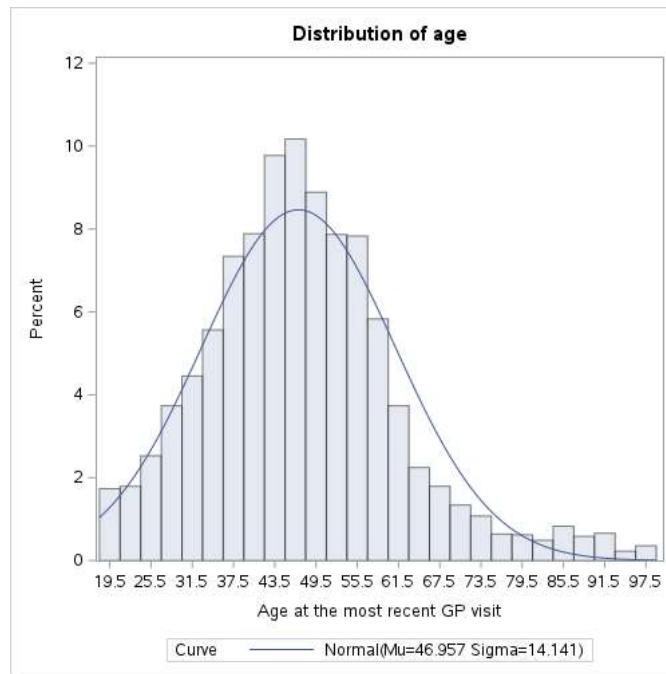
Being obese (BMI>=30)		
Obese_GP	Frequency	Percent
No	3607	69.25
Yes	1602	30.75
Frequency Missing = 91		

Have high blood pressure (>=135/85mmHg)		
HighBP_GP	Frequency	Percent
No	3650	68.87
Yes	1650	31.13

Reason for the most recent GP visit		
reason	Frequency	Percent
ABDOMINAL PAIN	241	4.55
DIZZINESS	275	5.19
HALLUCINATIONS	245	4.62
HEADACHE	2637	49.75
ITCHING	273	5.15
NAUSEA	594	11.21
PALPITATIONS	243	4.58
SKIN RASH	270	5.09

Reason for the most recent GP visit		
reason	Frequency	Percent
TINNITUS	275	5,19
VOMITING	247	4,66

The UNIVARIATE Procedure



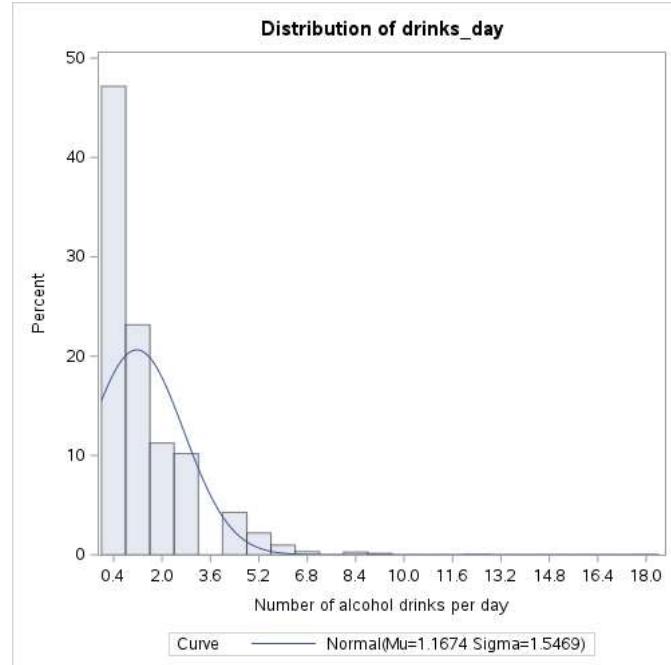
The UNIVARIATE Procedure
Fitted Normal Distribution for age (Age at the most recent GP visit)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	46.95717
Std Dev	Sigma	14.14148

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0,0569240	Pr > D	<0,010
Cramer-von Mises	W-Sq	3,9389774	Pr > W-Sq	<0,005
Anderson-Darling	A-Sq	29,4844122	Pr > A-Sq	<0,005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	19,0000	14,0592
5.0	25,0000	23,6965
10.0	30,0000	28,8341
25.0	38,0000	37,4189
50.0	46,0000	46,9572
75.0	55,0000	56,4955
90.0	64,0000	65,0802
95.0	73,0000	70,2178
99.0	90,5000	79,8552

The UNIVARIATE Procedure



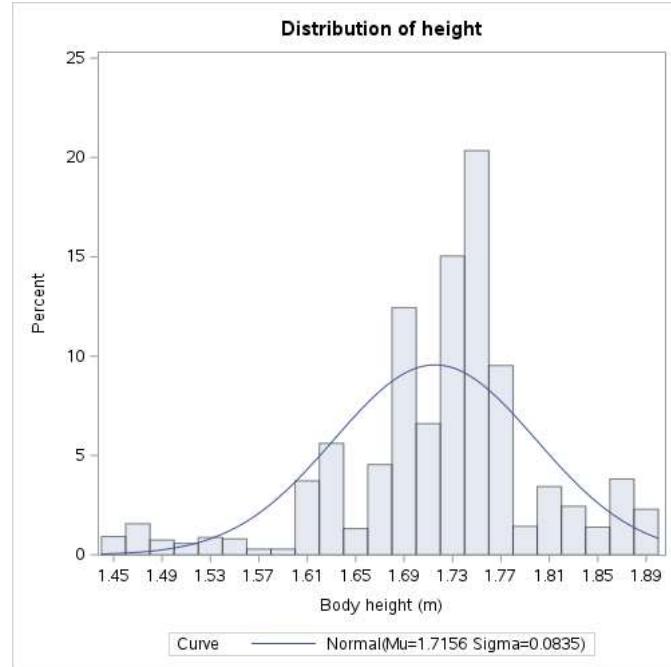
The UNIVARIATE Procedure
Fitted Normal Distribution for drinks_day (Number of alcohol drinks per day)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	μ	1.167399
Std Dev	σ	1.546852

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.246418	$Pr > D$ <0.010
Cramer-von Mises	W-Sq		$Pr > W-Sq$ <0.005
Anderson-Darling	A-Sq		$Pr > A-Sq$ <0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	0.00000	-2.43112
5.0	0.00000	-1.37695
10.0	0.00000	-0.81497
25.0	0.00000	0.12406
50.0	1.00000	1.16740
75.0	2.00000	2.21073
90.0	3.00000	3.14977
95.0	4.00000	3.71174
99.0	6.00000	4.76592

The UNIVARIATE Procedure



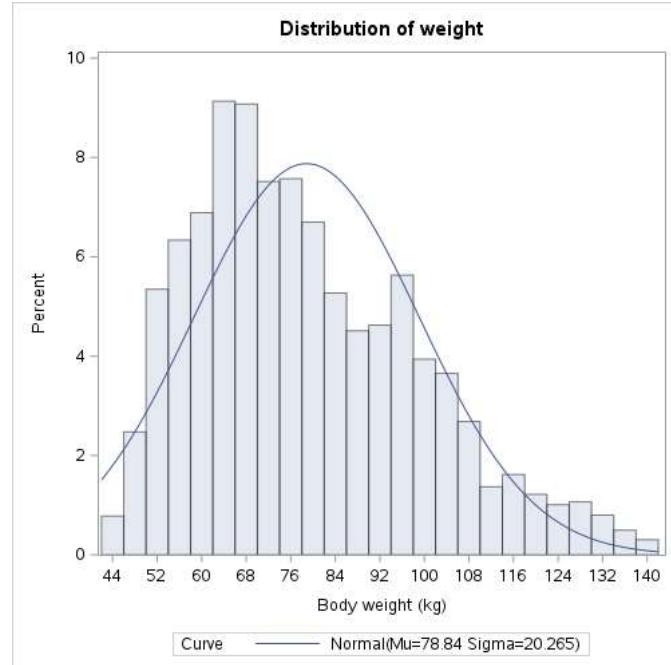
The UNIVARIATE Procedure
Fitted Normal Distribution for height (Body height (m))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	μ	1.715552
Std Dev	σ	0.083477

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.124895	$Pr > D$ <0.010
Cramer-von Mises	W-Sq		$Pr > W\text{-Sq}$ <0.005
Anderson-Darling	A-Sq		$Pr > A\text{-Sq}$ <0.005

Percent	Quantile	
	Observed	Estimated
1.0	1.46000	1.52136
5.0	1.54000	1.57824
10.0	1.62000	1.60857
25.0	1.68000	1.65925
50.0	1.73000	1.71555
75.0	1.75000	1.77186
90.0	1.81000	1.82253
95.0	1.86000	1.85286
99.0	1.88000	1.90975

The UNIVARIATE Procedure



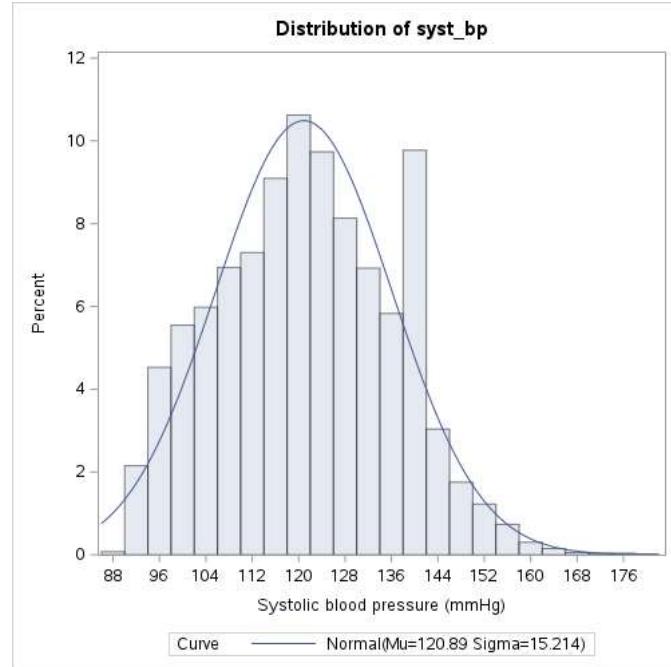
The UNIVARIATE Procedure
Fitted Normal Distribution for weight (Body weight (kg))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	78.8399
Std Dev	Sigma	20.26496

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.0736349	Pr > D <0.010
Cramer-von Mises	W-Sq	9.2512600	Pr > W-Sq <0.005
Anderson-Darling	A-Sq	56.6893041	Pr > A-Sq <0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	46.4000	31.6966
5.0	51.8000	45.5070
10.0	54.8000	52.8693
25.0	63.3000	65.1714
50.0	75.2000	78.8399
75.0	92.9000	92.5084
90.0	106.5000	104.8105
95.0	117.6000	112.1728
99.0	133.0000	125.9832

The UNIVARIATE Procedure



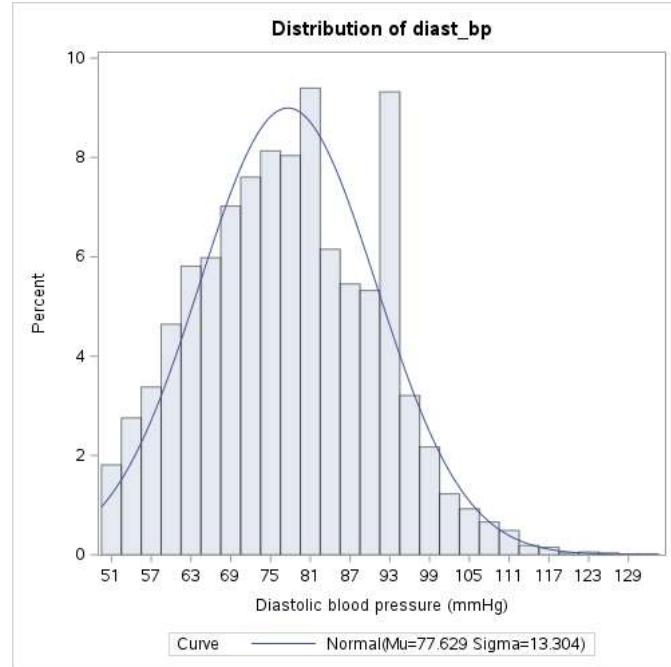
The UNIVARIATE Procedure
Fitted Normal Distribution for syst_bp (Systolic blood pressure (mmHg))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	μ	120.8864
Std Dev	σ	15.21383

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.0438037	Pr > D <0.010
Cramer-von Mises	W-Sq	1.7870507	Pr > W-Sq <0.005
Anderson-Darling	A-Sq	15.0513570	Pr > A-Sq <0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	92.0000	85.4938
5.0	96.0000	95.8619
10.0	100.0000	101.3891
25.0	109.0000	110.6248
50.0	120.0000	120.8864
75.0	132.0000	131.1480
90.0	140.0000	140.3837
95.0	144.5000	145.9109
99.0	155.0000	156.2791

The UNIVARIATE Procedure



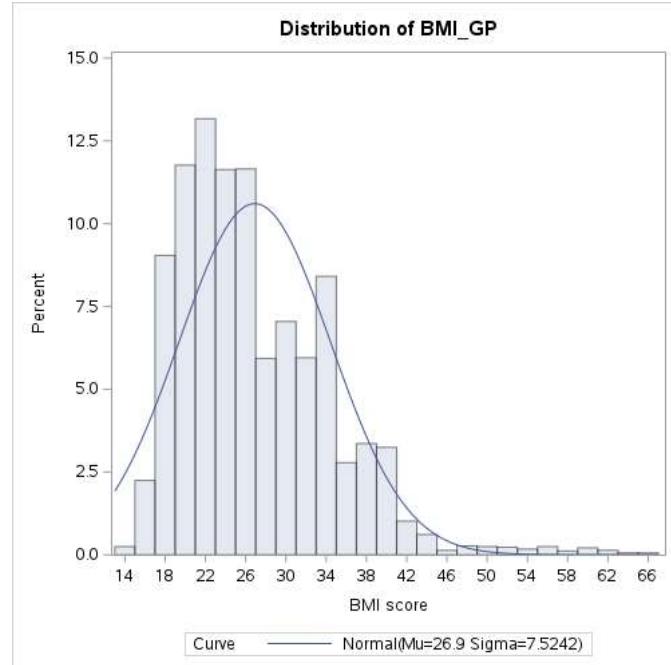
The UNIVARIATE Procedure
Fitted Normal Distribution for diast_bp (Diastolic blood pressure (mmHg))

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	μ	77.62887
Std Dev	σ	13.30412

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic	p Value	
Kolmogorov-Smirnov	D	0.0450681	<0.010
Cramer-von Mises	W-Sq	1.5854582	<0.005
Anderson-Darling	A-Sq	11.6936332	<0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	51.0000	46.6789
5.0	56.0000	55.7455
10.0	60.0000	60.5790
25.0	68.0000	68.6554
50.0	78.0000	77.6289
75.0	88.0000	86.6024
90.0	94.0000	94.6788
95.0	99.0000	99.5122
99.0	109.5000	108.5789

The UNIVARIATE Procedure

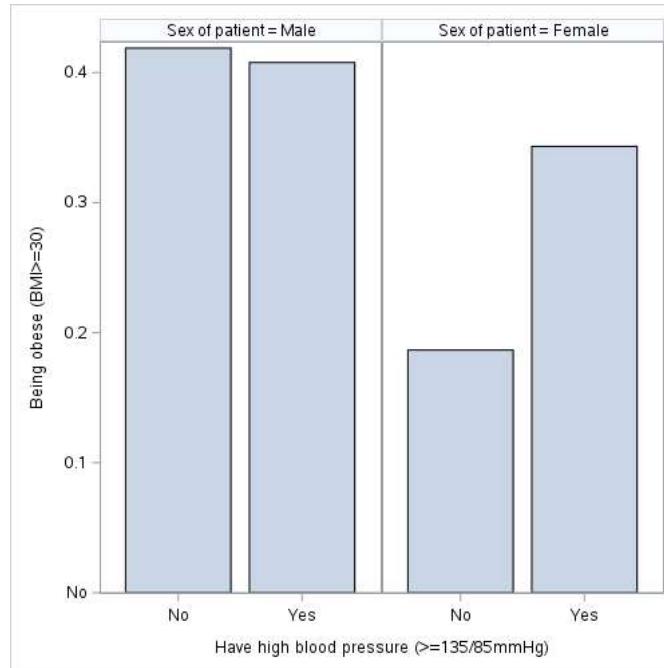


The UNIVARIATE Procedure
Fitted Normal Distribution for BMI_GP (BMI score)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	μ	26.90041
Std Dev	σ	7.52424

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0,0936832	Pr > D <0,010
Cramer-von Mises	W-Sq	13,8085810	Pr > W-Sq <0,005
Anderson-Darling	A-Sq	84,2381722	Pr > A-Sq <0,005

Percent	Quantile	
	Observed	Estimated
1.0	15,6504	9,39641
5.0	17,9151	14,52414
10.0	18,7444	17,25771
25.0	21,2585	21,82539
50.0	25,2092	26,90041
75.0	31,8754	31,97543
90.0	37,1173	36,54311
95.0	39,9660	39,27668
99.0	52,3400	44,40441



The CONTENTS Procedure

Data Set Name	WORK.ED	Observations	30466
Member Type	DATA	Variables	15
Engine	V9	Indexes	0
Created	21/02/2024 20:48:10	Observation Length	96
Last Modified	21/02/2024 20:48:10	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	23
First Data Page	1
Max Obs per Page	1363
Obs in First Data Page	1319
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A680001DAF4_odaws02-apse1-2.oda.sas.com/SAS_workC01E0001DAF4_odaws02-apse1-2.oda.sas.com/ed.sas7bd1
Release Created	9.0401M7
Host Created	Linux
Inode Number	525845
Access Permission	rw-r--r--
Owner Name	u62298616
File Size	3MB
File Size (bytes)	3145728

Variables in Creation Order						
#	Variable	Type	Len	Format	Informat	Label
1	ID	Num	8			Unique person ID
2	ed_admission	Num	8	DDMMYY10.		Date of ED attendance
3	ed_separation	Num	8	DDMMYY10.		Date of ED separation
4	age_ed	Num	8			Age of patient at ED attendance
5	sex_ed	Num	8	SEXF.		Sex of patient
6	cob_ed	Num	8	COBF.		Country of birth
7	interpreter	Num	8	YNF.		Interpreter required
8	health_insurance	Num	8	YNF.		Have private health insurance
9	triage_category	Num	8	TRIAGEF.		Urgency of presentation
10	dx1	Char	3	\$3.	\$3.	Principal presenting diagnosis (ICD-10-AM codes)
11	dx2	Char	3	\$3.	\$3.	Additional presenting diagnosis 1 (ICD-10-AM codes)
12	dx3	Char	3	\$3.	\$3.	Additional presenting diagnosis 2 (ICD-10-AM codes)
13	dx4	Char	3	\$3.	\$3.	Additional presenting diagnosis 3 (ICD-10-AM codes)
14	dx5	Char	3	\$3.	\$3.	Additional presenting diagnosis 4 (ICD-10-AM codes)
15	separation mode	Num	8	SEPMODEF.		Status at separation from ED

Obs	ID	ed_admission	ed_separation	age_ed	sex_ed	cob_ed	interpreter	health_insurance	triage_category	dx1	dx2	dx3	dx4	dx5	separation_mode
-----	----	--------------	---------------	--------	--------	--------	-------------	------------------	-----------------	-----	-----	-----	-----	-----	-----------------

Obs	ID	ed_admission	ed_separation	age_ed	sex_ed	cob_ed	interpreter	health_insurance	triage_category	dx1	dx2	dx3	dx4	dx5	separation_mode
1	1	16/06/2014	21/06/2014	59	Female	Overseas	No	Yes	Urgent	F10	Z50	H82	N93		Departed ED
2	5	19/04/2014	21/04/2014	57	Female	Australia	No	No	Non urgent	S69	Z03	N12	I31	R04	Departed ED
3	5	16/06/2014	21/06/2014	57	Female	Australia	No	No	Urgent	S05	B23	M53	Z49	Z49	Departed ED
4	5	02/07/2014	03/07/2014	57	Female	Australia	No	No	Urgent	R33	Z48	Z46	Z48	O60	Departed ED
5	5	17/07/2014	18/07/2014	57	Female	Australia	No	No	Non urgent	R60	K40	Z39	J38	N44	Departed ED
6	5	30/07/2014	31/07/2014	57	Female	Australia	No	No	Semi urgent	Z48	Z40	M02	S09	Z39	Admitted to hospital
7	6	11/10/2014	12/10/2014	56	Female	Australia	No	No	Urgent	H27	S74	R93	Z40	Z11	Departed ED
8	6	23/10/2014	24/10/2014	56	Female	Australia	No	No	Urgent	Z39	R93	S66	O74	J94	Departed ED
9	6	30/10/2014	30/10/2014	56	Female	Australia	No	No	Semi urgent	M02	I34	A40	Z50	R68	Admitted to hospital
10	6	15/12/2014	19/12/2014	56	Female	Australia	No	No	Urgent	K92	B96	M91	H13	P13	Admitted to hospital

The MEANS Procedure

Variable	Label	Mean	Median	Mode	Std Dev	Variance	Minimum	Maximum	N Miss
ID	Unique person ID	10659.18	12002.00	10760.00	4355.48	18970171.75	1.00	15588.00	0
age_ed	Age of patient at ED attendance	49.04	47.00	47.00	19.43	377.59	18.00	111.00	0
ed_admission	Date of ED attendance	19917.36	19922.00	19965.00	99.91	9981.10	19724.00	20088.00	0
ed_separation	Date of ED separation	19919.19	19924.00	19941.00	99.94	9987.24	19724.00	20088.00	0

The FREQ Procedure

Sex of patient		
sex_ed	Frequency	Percent
Male	15279	50.49
Female	14981	49.51
Frequency Missing = 206		

Country of birth		
cob_ed	Frequency	Percent
Australia	21223	72.19
Overseas	8176	27.81
Frequency Missing = 1067		

Interpreter required		
interpreter	Frequency	Percent
No	24365	80.69
Yes	5829	19.31
Frequency Missing = 272		

Have private health insurance		
health_insurance	Frequency	Percent
No	13440	45.50
Yes	16097	54.50
Frequency Missing = 929		

Urgency of presentation		
triage_category	Frequency	Percent
Resuscitation	607	1.99
Emergency	1083	3.55
Urgent	9064	29.75
Semi urgent	13328	43.75
Non urgent	6384	20.95

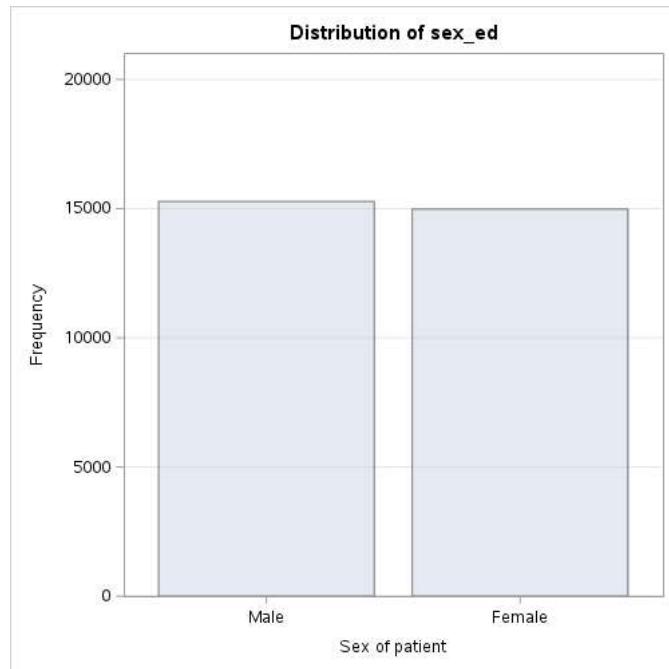
Status at separation from ED		
separation_mode	Frequency	Percent
Admitted to hospital	10907	35.80
Departed ED	19387	63.63
Died in ED	124	0.41
Dead on arrival	48	0.16

dx1	dx2	dx3	dx4	dx5
0	0	0	0	0

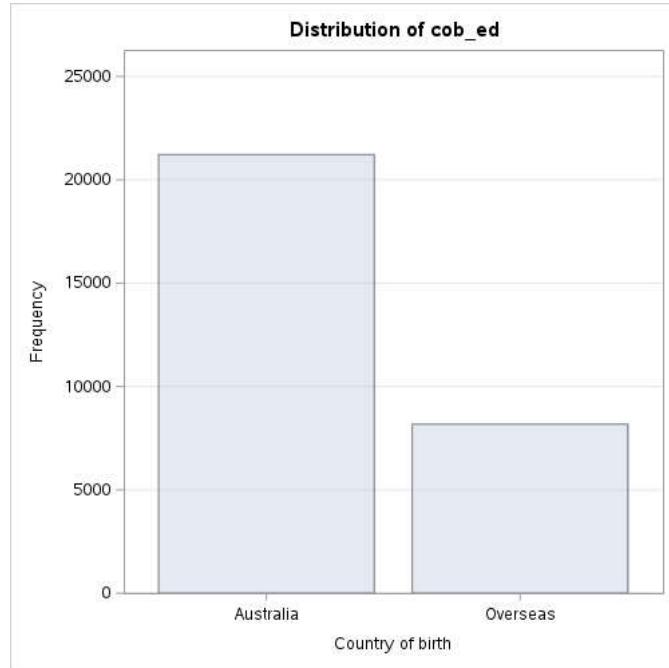
The FREQ Procedure

Sex of patient				
sex_ed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	15279	50.49	15279	50.49
Frequency Missing = 206				

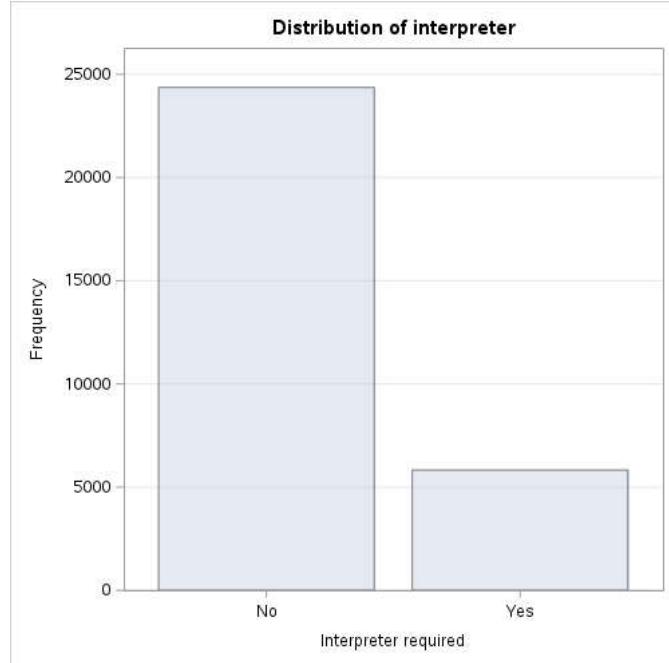
Sex of patient				
sex_ed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	14981	49.51	30260	100.00
Frequency Missing = 206				



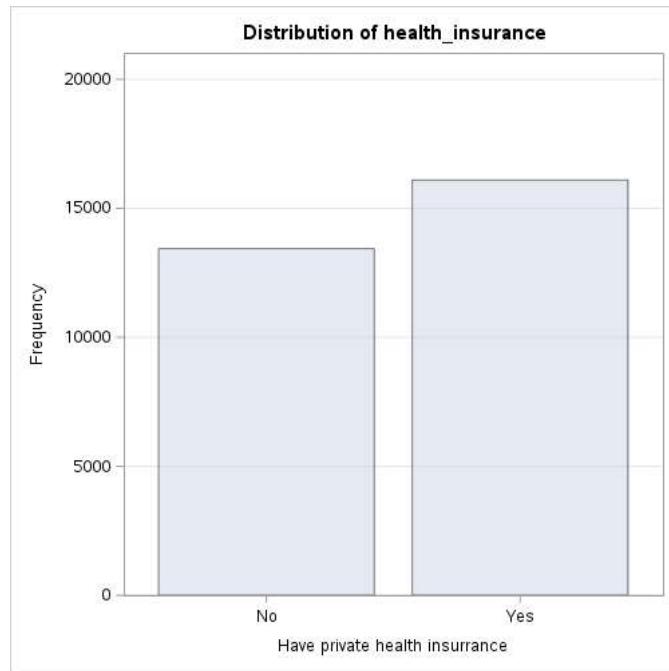
Country of birth				
cob_ed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Australia	21223	72.19	21223	72.19
Overseas	8176	27.81	29399	100.00
Frequency Missing = 1067				



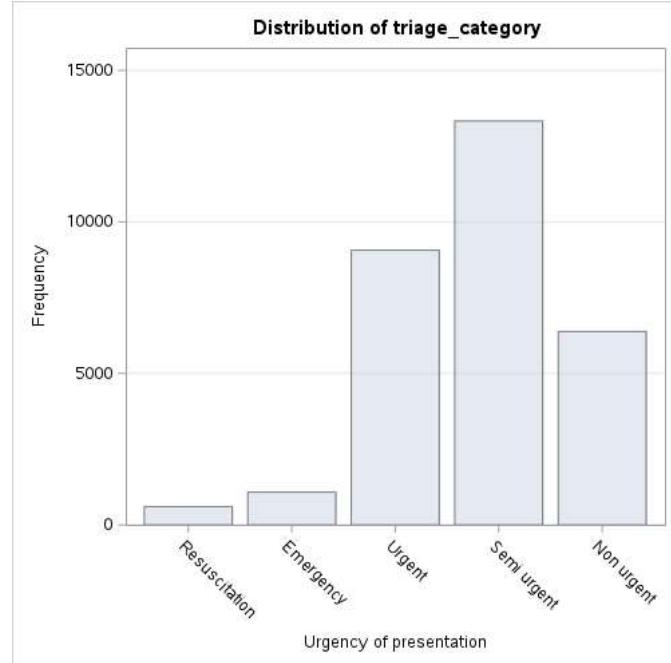
Interpreter required				
interpreter	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	24365	80.69	24365	80.69
Yes	5829	19.31	30194	100.00
Frequency Missing = 272				



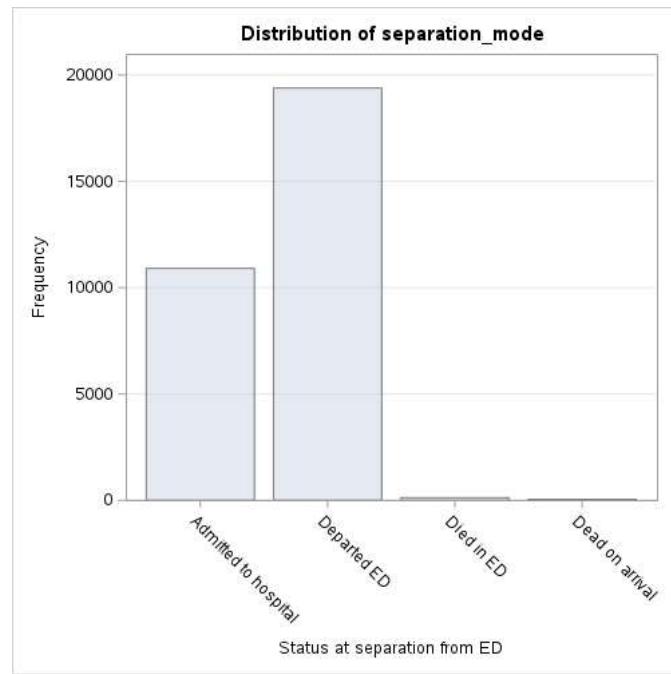
Have private health insurance				
health_insurance	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	13440	45.50	13440	45.50
Yes	16097	54.50	29537	100.00
Frequency Missing = 929				

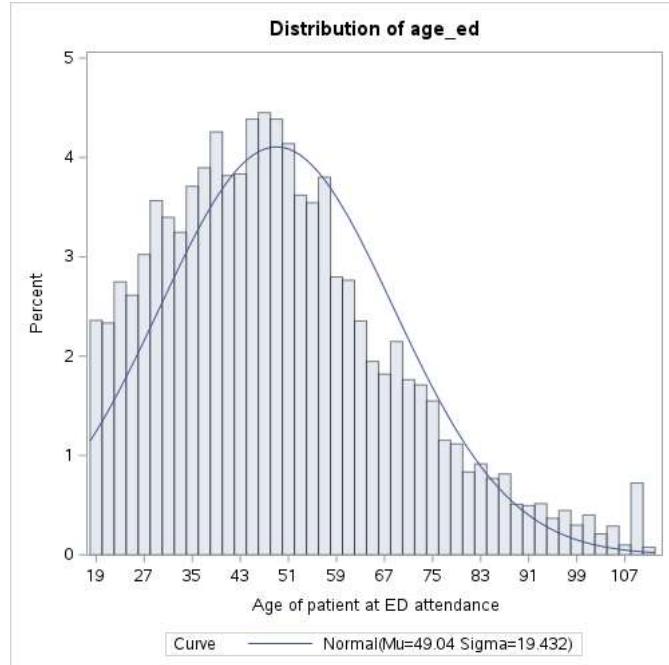


Urgency of presentation				
triage_category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Resuscitation	607	1.99	607	1.99
Emergency	1083	3.55	1690	5.55
Urgent	9064	29.75	10754	35.30
Semi urgent	13328	43.75	24082	79.05
Non urgent	6384	20.95	30466	100.00



Status at separation from ED				
separation_mode	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Admitted to hospital	10907	35.80	10907	35.80
Departed ED	19387	63.63	30294	99.44
Died in ED	124	0.41	30418	99.84
Dead on arrival	48	0.16	30466	100.00





The UNIVARIATE Procedure
Fitted Normal Distribution for age_ed (Age of patient at ED attendance)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	49.04018
Std Dev	Sigma	19.43177

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.061421	Pr > D <0.010
Cramer-von Mises	W-Sq		Pr > W-Sq <0.005
Anderson-Darling	A-Sq		Pr > A-Sq <0.005

Percent	Quantile	
	Observed	Estimated
1.0	18.0000	3.83513
5.0	22.0000	17.07777
10.0	25.0000	24.13737
25.0	34.0000	35.93365
50.0	47.0000	49.04018
75.0	60.0000	62.14670
90.0	76.0000	73.94299
95.0	86.0000	81.00259
99.0	105.0000	94.24522

The CONTENTS Procedure

Data Set Name	WORK.PBS	Observations	3164
Member Type	DATA	Variables	6
Engine	V9	Indexes	0
Created	21/02/2024 20:48:10	Observation Length	688
Last Modified	21/02/2024 20:48:10	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	17
First Data Page	1
Max Obs per Page	190
Obs in First Data Page	186
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A680001DAF4_odaws02-apse1-2.oda,sas.com/SAS_workC01E0001DAF4_odaws02-apse1-2.oda,sas.com/pbs,sas7bdat
Release Created	9.0401M7

Engine/Host Dependent Information	
Host Created	Linux
Inode Number	525849
Access Permission	rwx-r--r--
Owner Name	u62298616
File Size	2MB
File Size (bytes)	2359296

Variables in Creation Order						
#	Variable	Type	Len	Format	Informat	Label
1	ID	Num	8			Unique person ID
2	supply_date	Num	8	DDMMYY10.	DDMMYY10.	Date of medication dispensed
3	ATC	Char	7	\$7.	\$7.	Anatomical Therapeutic Chemical code
4	drug_name	Char	250	\$250.	\$250.	Generic name of medicine
5	item_code	Char	10	\$10.	\$10.	PBS item code
6	form_strength	Char	400	\$400.	\$400.	Form and strength of medicine

Obs	ID	supply_date	ATC	drug_name	item_code	form_strength
1	15	19/01/2014	N07BA03	VARENICLINE	09129L	Tablet 1mg (as tartrate) 112
2	15	27/03/2014	A02BC02	PANTOPRAZOLE	09424B	Sachet containing granules 40mg 30
3	15	16/07/2014	A07EC01	SULPHASALAZINE	09209Q	Tablet 500mg (enteric coated) 200
4	15	07/08/2014	A07EC01	SULPHASALAZINE	09209Q	Tablet 500mg (enteric coated) 200
5	15	22/09/2014	A02BC02	PANTOPRAZOLE	09424B	Sachet containing granules 40mg 30
6	41	29/04/2014	N05AH04	QUETIAPINE	09205L	Tablet (modified release) equivalent to 400mg 60
7	41	24/08/2014	N05AH04	QUETIAPINE	09205L	Tablet (modified release) equivalent to 400mg 60
8	41	24/08/2014	N07BA03	VARENICLINE	09129L	Tablet 1mg (as tartrate) 112
9	41	28/11/2014	N05AH04	QUETIAPINE	09205L	Tablet (modified release) equivalent to 400mg 60
10	100	28/02/2014	N07BA03	VARENICLINE	09129L	Tablet 1mg (as tartrate) 112

The MEANS Procedure

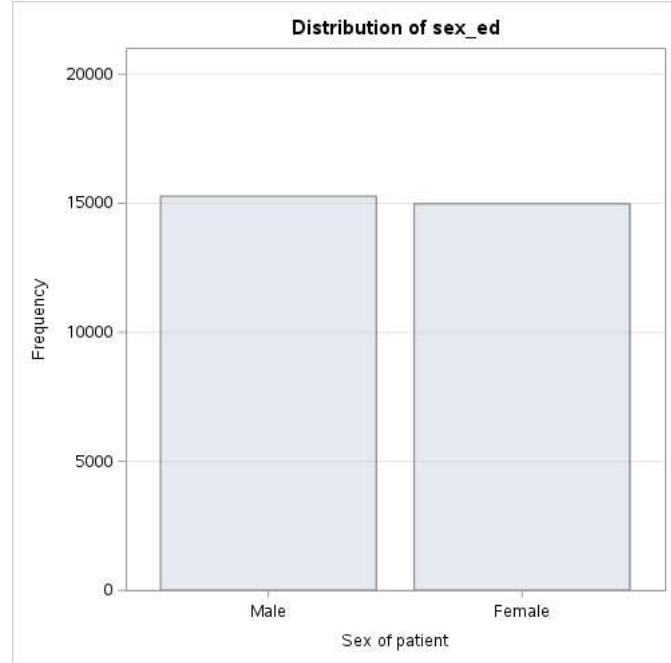
Variable	Label	Mean	Median	Mode	Std Dev	Variance	Minimum	Maximum	N Miss
ID	Unique person ID	6186.81	4763.00	13465.00	4250.29	18065006.68	15.00	15580.00	0
supply_date	Date of medication dispensed	19904.61	19906.00	19968.00	104.23	10864.78	19724.00	20088.00	0

ATC	drug_name	item_code	form_strength
247	244	247	225

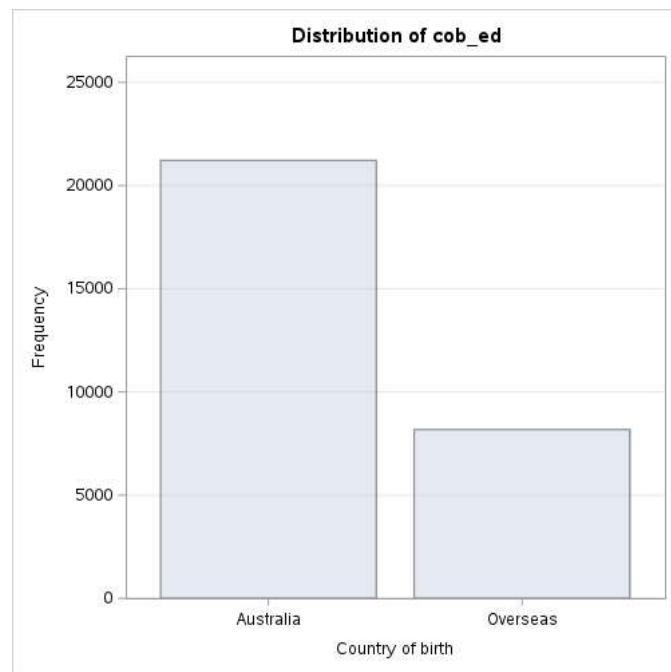
ATC	drug_name	item_code	form_strength
0	0	0	0

The FREQ Procedure

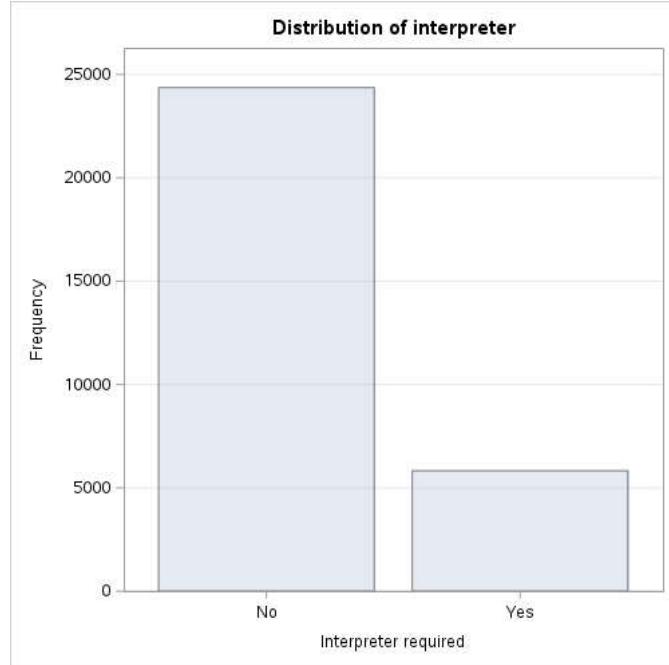
Sex of patient				
sex_ed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	15279	50.49	15279	50.49
Female	14981	49.51	30260	100.00
Frequency Missing = 206				



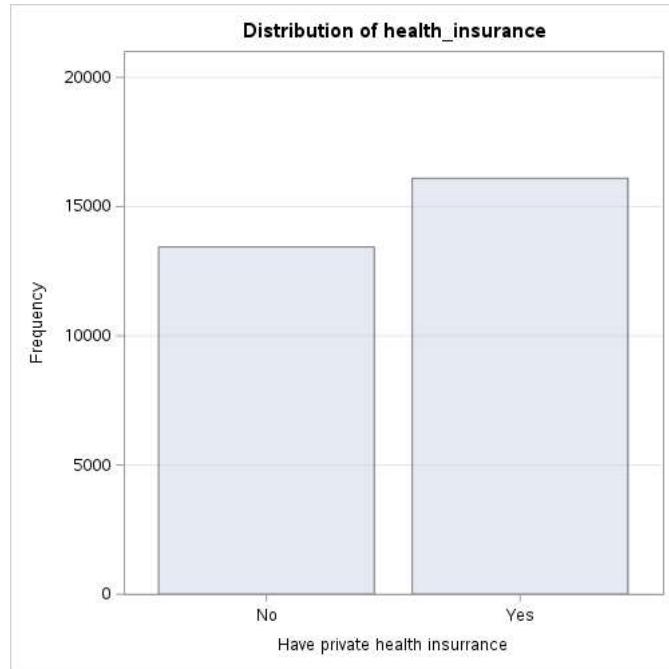
Country of birth				
cob_ed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Australia	21223	72.19	21223	72.19
Overseas	8176	27.81	29399	100.00
Frequency Missing = 1067				



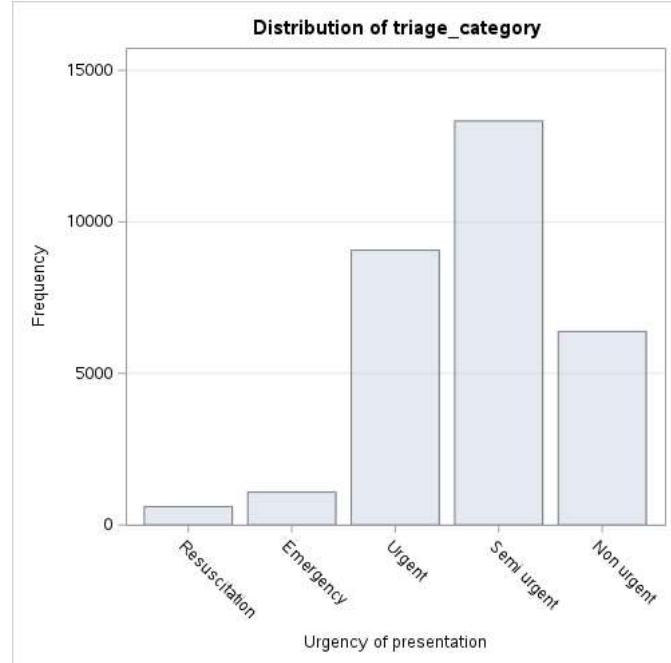
Interpreter required				
interpreter	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	24365	80.69	24365	80.69
Yes	5829	19.31	30194	100.00
Frequency Missing = 272				



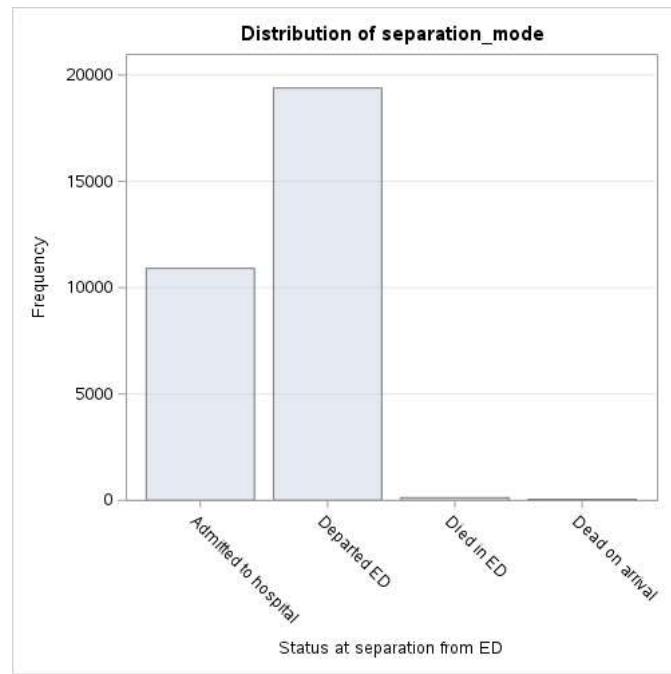
Have private health insurance				
health_insurance	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	13440	45.50	13440	45.50
Yes	16097	54.50	29537	100.00
Frequency Missing = 929				

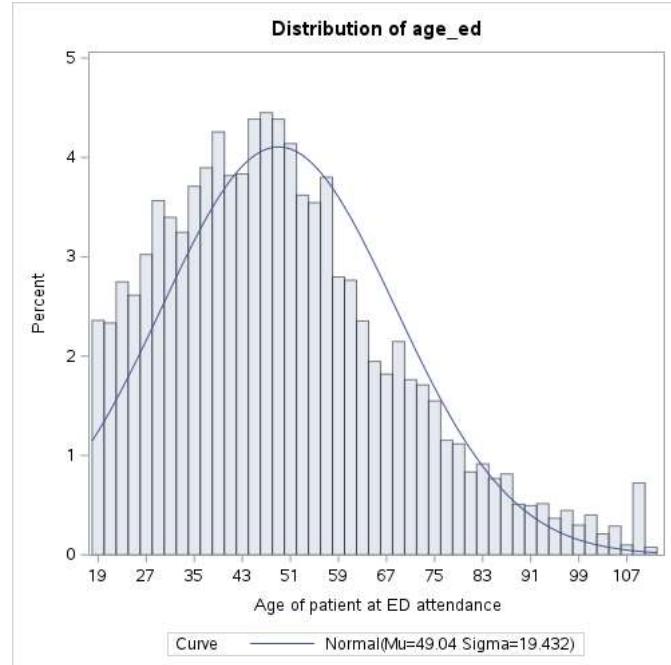


Urgency of presentation				
triage_category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Resuscitation	607	1.99	607	1.99
Emergency	1083	3.55	1690	5.55
Urgent	9064	29.75	10754	35.30
Semi urgent	13328	43.75	24082	79.05
Non urgent	6384	20.95	30466	100.00



Status at separation from ED				
separation_mode	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Admitted to hospital	10907	35.80	10907	35.80
Departed ED	19387	63.63	30294	99.44
Died in ED	124	0.41	30418	99.84
Dead on arrival	48	0.16	30466	100.00





The UNIVARIATE Procedure
Fitted Normal Distribution for age_ed (Age of patient at ED attendance)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	49.04018
Std Dev	Sigma	19.43177

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.061421	Pr > D <0.010
Cramer-von Mises	W-Sq		Pr > W-Sq <0.005
Anderson-Darling	A-Sq		Pr > A-Sq <0.005

Quantiles for Normal Distribution		
Percent	Quantile	
	Observed	Estimated
1.0	18.0000	3.83513
5.0	22.0000	17.07777
10.0	25.0000	24.13737
25.0	34.0000	35.93365
50.0	47.0000	49.04018
75.0	60.0000	62.14670
90.0	76.0000	73.94299
95.0	86.0000	81.00259
99.0	105.0000	94.24522

The FREQ Procedure

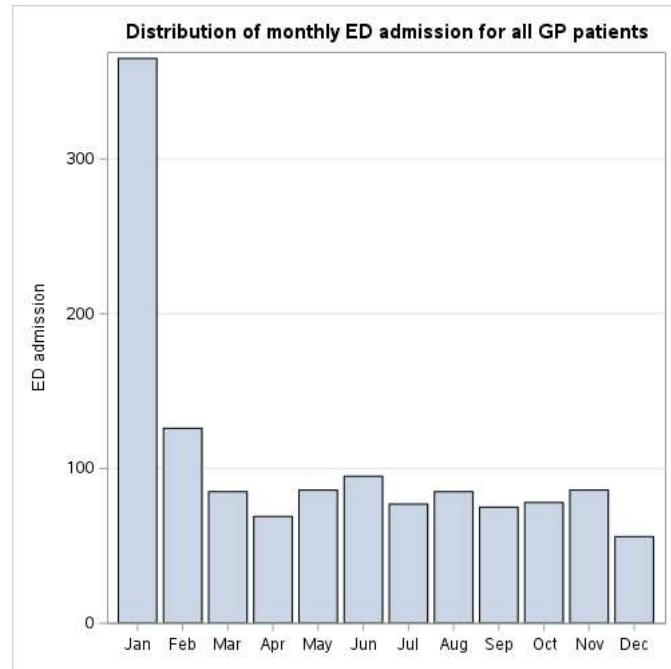
Agegroup_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Under 60 years old	4526	85.40	4526	85.40
60 years old and older	774	14.60	5300	100.00

The FREQ Procedure

ed_record	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4017	75.79	4017	75.79
Yes	1283	24.21	5300	100.00

Total_number
1283

Total_number
4017



Distribution of monthly ED admission for all GP patients

The FREQ Procedure

Frequency Col Pct	Table of sex by ed_record		
	ed_record		
	No	Yes	Total
Male	1720 42.91	644 50.23	2364
Female	2288 57.09	638 49.77	2926
Total	4008	1282	5290

Frequency Missing = 10

Statistics for Table of sex by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	21.0549	<.0001
Likelihood Ratio Chi-Square	1	20.9760	<.0001
Continuity Adj. Chi-Square	1	20.7598	<.0001
Mantel-Haenszel Chi-Square	1	21.0509	<.0001
Phi Coefficient		-0.0631	
Contingency Coefficient		0.0630	
Cramer's V		-0.0631	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	1720
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1,0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5290
Frequency Missing = 10

Frequency Col Pct	Table of Agegroup_GP by ed_record		
	ed_record		
	No	Yes	Total
Under 60 years old	3435 85.51	1091 85.04	4526
60 years old and older	582 14.49	192 14.96	774
Total	4017	1283	5300

Statistics for Table of Agegroup_GP by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	0.1770	0.6739
Likelihood Ratio Chi-Square	1	0.1762	0.6746
Continuity Adj. Chi-Square	1	0.1409	0.7074
Mantel-Haenszel Chi-Square	1	0.1770	0.6740
Phi Coefficient		0.0058	
Contingency Coefficient		0.0058	
Cramer's V		0.0058	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3435
Left-sided Pr <= F	0.6810
Right-sided Pr >= F	0.3519
Table Probability (P)	0.0329
Two-sided Pr <= P	0.6829

Sample Size = 5300

Frequency Col Pct	Table of cob by ed_record		
	ed_record		
cob(Country of birth)	No	Yes	Total
Australia	1647 41.65	586 45.67	2233
Overseas	2307 58.35	697 54.33	3004
Total	3954	1283	5237

Frequency Missing = 63

Statistics for Table of cob by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	6.4010	0.0114
Likelihood Ratio Chi-Square	1	6.3780	0.0116
Continuity Adj. Chi-Square	1	6.2377	0.0125
Mantel-Haenszel Chi-Square	1	6.3998	0.0114
Phi Coefficient		-0.0350	
Contingency Coefficient		0.0349	
Cramer's V		-0.0350	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	1647
Left-sided Pr <= F	0.0063
Right-sided Pr >= F	0.9947
Table Probability (P)	0.0011
Two-sided Pr <= P	0.0124

Sample Size = 5237
Frequency Missing = 63

Frequency Col Pct	Table of healthcare_card by ed_record		
	ed_record		
healthcare_card(Have a health care card)	No	Yes	Total
No	2903 72.36	701 54.77	3604
Yes	1109 27.64	579 45.23	1688
Total	4012	1280	5292

Frequency Missing = 8

Statistics for Table of healthcare_card by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	138.2544	<.0001
Likelihood Ratio Chi-Square	1	133.2717	<.0001
Continuity Adj. Chi-Square	1	137.4457	<.0001
Mantel-Haenszel Chi-Square	1	138.2282	<.0001
Phi Coefficient		0.1616	
Contingency Coefficient		0.1596	
Cramer's V		0.1616	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2903
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001

Fisher's Exact Test		
Two-sided Pr <= P	<.0001	

Sample Size = 5292
Frequency Missing = 8

Frequency
Col Pct

Table of Smoke_current_GP by ed_record			
Smoke_current_GP(Being a current smoker)	ed_record		
	No	Yes	Total
No	3481 87.18	1014 80.54	4495
Yes	512 12.82	245 19.46	757
Total	3993	1259	5252

Frequency Missing = 48

Statistics for Table of Smoke_current_GP by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	34.1843	<.0001
Likelihood Ratio Chi-Square	1	32.3072	<.0001
Continuity Adj. Chi-Square	1	33.6484	<.0001
Mantel-Haenszel Chi-Square	1	34.1778	<.0001
Phi Coefficient		0.0807	
Contingency Coefficient		0.0804	
Cramer's V		0.0807	

Fisher's Exact Test			
Cell (1,1) Frequency (F)	3481		
Left-sided Pr <= F	1.0000		
Right-sided Pr >= F	<.0001		
Table Probability (P)	<.0001		
Two-sided Pr <= P	<.0001		

Sample Size = 5252
Frequency Missing = 48

Frequency
Col Pct

Table of Risky_alcohol_GP by ed_record			
Risky_alcohol_GP(Have two or more alcohol drinks per day)	ed_record		
	No	Yes	Total
No	3304 87.22	779 63.96	4083
Yes	484 12.78	439 36.04	923
Total	3788	1218	5006

Frequency Missing = 294

Statistics for Table of Risky_alcohol_GP by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	331.7354	<.0001
Likelihood Ratio Chi-Square	1	298.0875	<.0001
Continuity Adj. Chi-Square	1	330.1902	<.0001
Mantel-Haenszel Chi-Square	1	331.6692	<.0001
Phi Coefficient		0.2574	
Contingency Coefficient		0.2493	
Cramer's V		0.2574	

Fisher's Exact Test			
Cell (1,1) Frequency (F)	3304		
Left-sided Pr <= F	1.0000		
Right-sided Pr >= F	<.0001		
Table Probability (P)	<.0001		
Two-sided Pr <= P	<.0001		

Sample Size = 5006
Frequency Missing = 294

Frequency
Col Pct

Table of Obese_GP by ed_record			
Obese_GP(Being obese (BMI>=30))	ed_record		
	No	Yes	Total
No	2907 73.59	700 55.60	3607
Yes	1043 26.41	559 44.40	1602
Total	3950	1259	5209

Frequency Missing = 91

Statistics for Table of Obese_GP by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	145.1726	<.0001
Likelihood Ratio Chi-Square	1	139.3714	<.0001
Continuity Adj. Chi-Square	1	144.3288	<.0001
Mantel-Haenszel Chi-Square	1	145.1447	<.0001
Phi Coefficient		0.1669	
Contingency Coefficient		0.1647	
Cramer's V		0.1669	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2907
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5209
Frequency Missing = 91

Frequency
Col Pct

		Table of HighBP_GP by ed_record		
		ed_record		
		No	Yes	Total
HighBP_GP(Have high blood pressure (>=135/85mmHg))	No	2915 72,57	735 57,29	3650
Yes		1102 27,43	548 42,71	1650
Total		4017	1283	5300

Statistics for Table of HighBP_GP by ed_record

Statistic	DF	Value	Prob
Chi-Square	1	105.8806	<.0001
Likelihood Ratio Chi-Square	1	102.2101	<.0001
Continuity Adj. Chi-Square	1	105.1692	<.0001
Mantel-Haenszel Chi-Square	1	105.8607	<.0001
Phi Coefficient		0.1413	
Contingency Coefficient		0.1400	
Cramer's V		0.1413	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2915
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5300

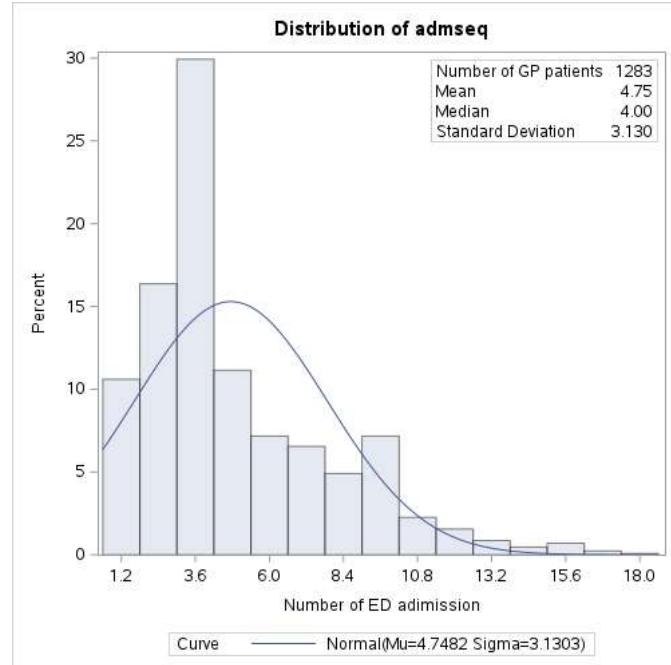
Distribution of monthly ED admission for all GP patients

The MEANS Procedure

Variable	Label	Minimum	Median	Mean	Maximum	Lower Quartile	Upper Quartile	10th Pctl	25th Pctl	50th Pctl	75th Pctl	90th Pctl	99th Pctl
ID admseq	Unique patientID	1,0000000 1,0000000	2831.00 4,0000000	2879.25 4,7482463	5793.00 18,0000000	1433.00 2,0000000	4384.00 6,0000000	556.0000000 1,0000000	1433.00 2,0000000	2831.00 4,0000000	4384.00 6,0000000	5190.00 9,0000000	5740.00 15,0000000

Distribution of number of ED admission per single GP patient

The UNIVARIATE Procedure



Distribution of number of ED admission per single GP patient

The UNIVARIATE Procedure
Fitted Normal Distribution for admseq (Number of ED admission)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	4.748246
Std Dev	Sigma	3.130256

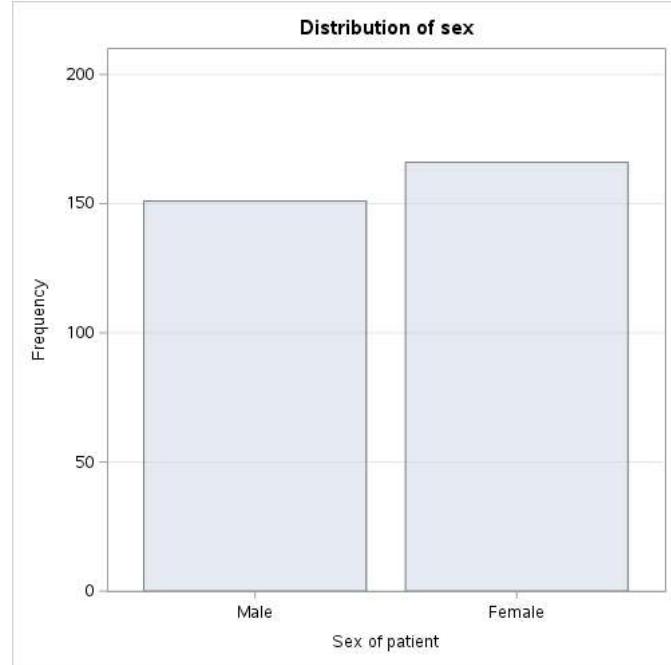
Goodness-of-Fit Tests for Normal Distribution				
Test		Statistic	p Value	
Kolmogorov-Smirnov	D	0.1634404	Pr > D	<0.010
Cramer-von Mises	W-Sq	6.3507912	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	37.4978861	Pr > A-Sq	<0.005

Percent	Quantile	
	Observed	Estimated
1.0	1.00000	-2.53382
5.0	1.00000	-0.40057
10.0	1.00000	0.73666
25.0	2.00000	2.63692
50.0	4.00000	4.74825
75.0	6.00000	6.85957
90.0	9.00000	8.75983
95.0	11.00000	9.89706
99.0	15.00000	12.03031

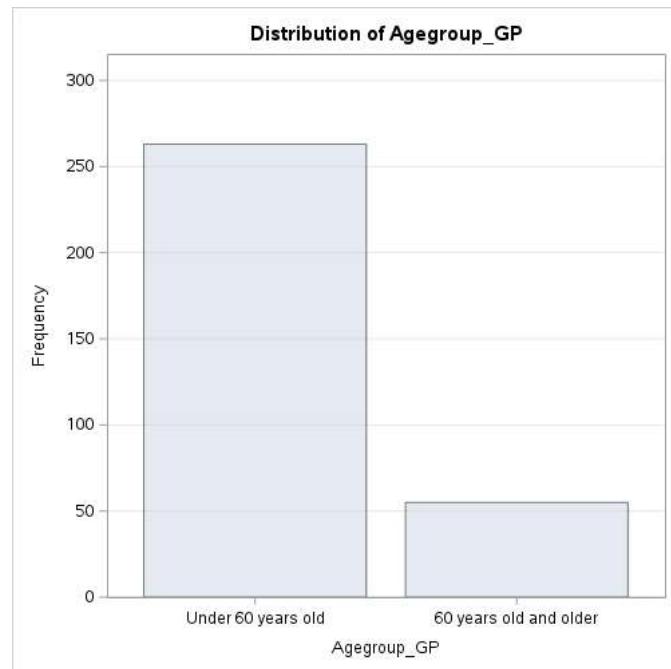
Socio-demographic report of the highest 25% of ED visit patients

The FREQ Procedure

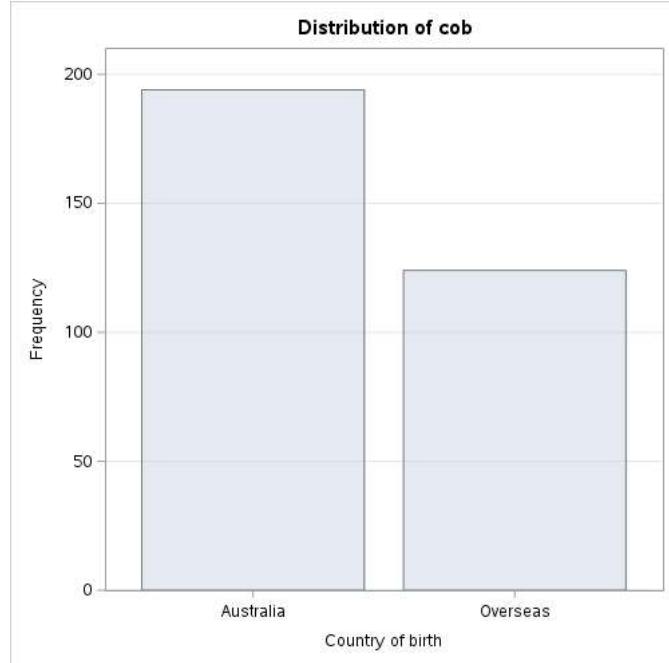
Sex of patient				
sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	151	47.63	151	47.63
Female	166	52.37	317	100.00
Frequency Missing = 1				



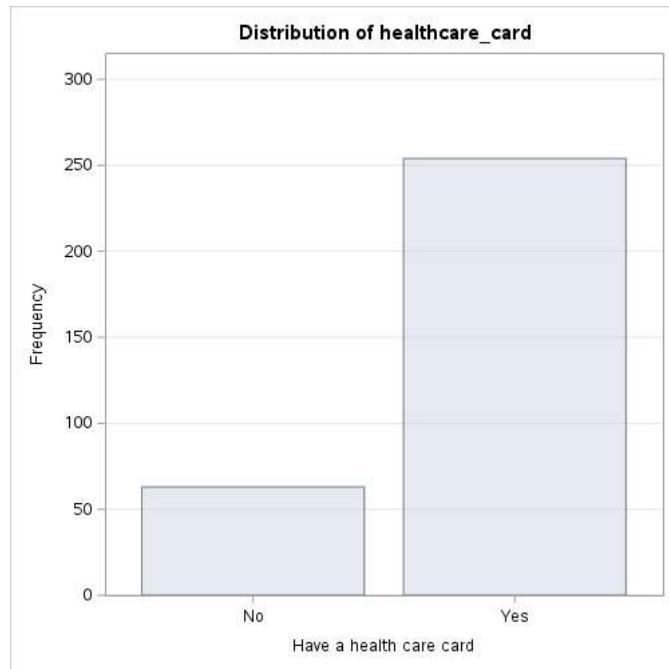
Agegroup_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Under 60 years old	263	82.70	263	82.70
60 years old and older	55	17.30	318	100.00



Country of birth				
cob	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Australia	194	61.01	194	61.01
Overseas	124	38.99	318	100.00



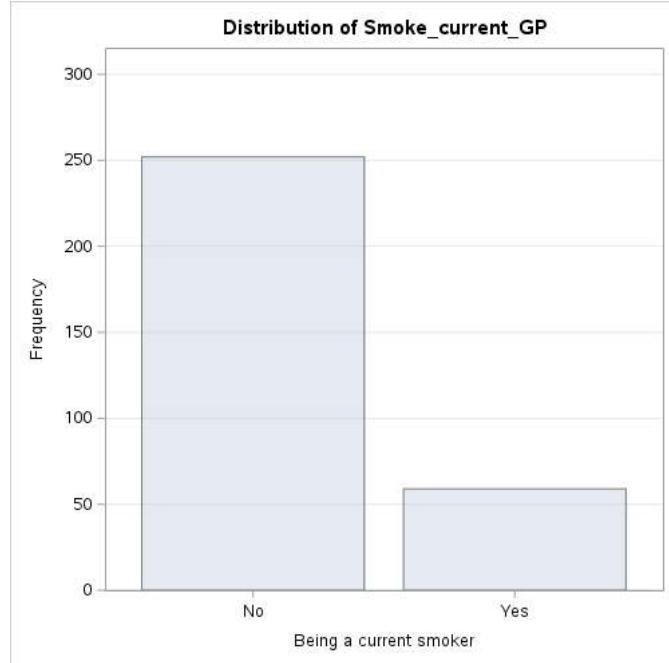
Have a health care card				
healthcare_card	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	63	19.87	63	19.87
Yes	254	80.13	317	100.00
Frequency Missing = 1				



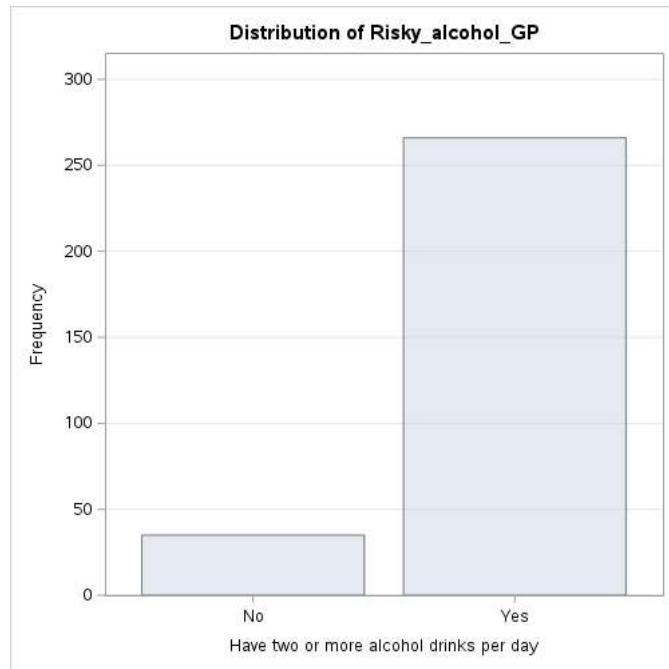
Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

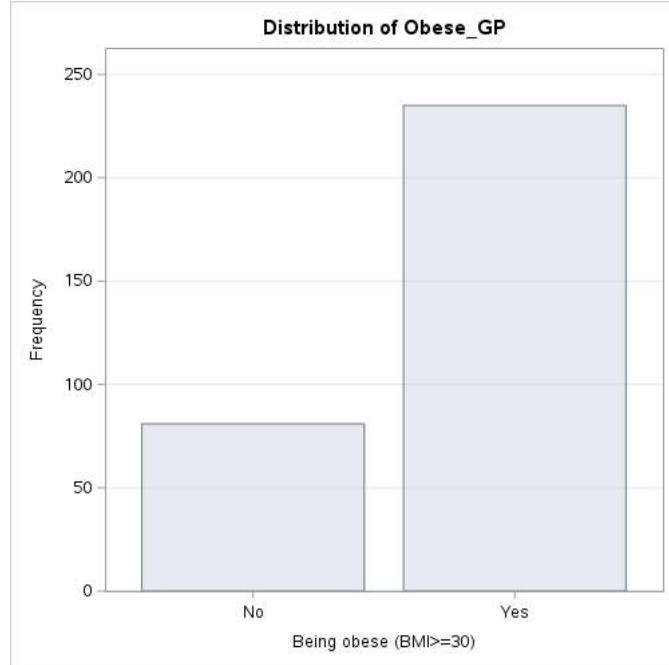
Being a current smoker				
Smoke_current_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	252	81.03	252	81.03
Yes	59	18.97	311	100.00
Frequency Missing = 7				



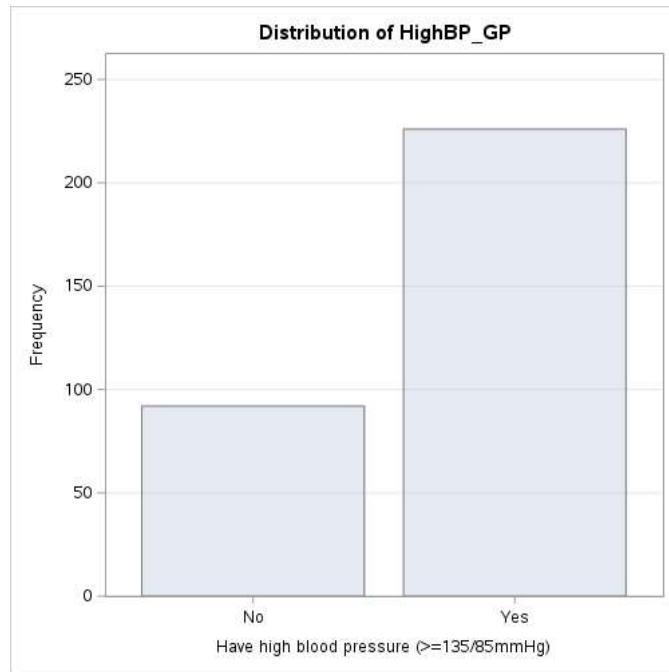
Have two or more alcohol drinks per day				
Risky_alcohol_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	35	11.63	35	11.63
Yes	266	88.37	301	100.00
Frequency Missing = 17				



Being obese (BMI>=30)				
Obese_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	81	25.63	81	25.63
Yes	235	74.37	316	100.00
Frequency Missing = 2				



Have high blood pressure (>=135/85mmHg)				
HighBP_GP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	92	28.93	92	28.93
Yes	226	71.07	318	100.00



Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

smoker_ED	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4758	84.41	4758	84.41
Yes, smoker	879	15.59	5637	100.00

risky_alcohol_ED	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4507	79.95	4507	79.95
Yes, drinker	1130	20.05	5637	100.00

obesity_ED	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4117	73.04	4117	73.04
Yes, obese	1520	26.96	5637	100.00

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

gp_record	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4354	77.24	4354	77.24
Yes	1283	22.76	5637	100.00

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Frequency Col Pct	Table of sex_ed by gp_record		
	gp_record		
sex_ed(Sex of patient)	No	Yes	Total
Male	2206 51.10	644 50.23	2850
Female	2111 48.90	638 49.77	2749
Total	4317	1282	5599

Frequency Missing = 38

Statistics for Table of sex_ed by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	0.2968	0.5859
Likelihood Ratio Chi-Square	1	0.2968	0.5859
Continuity Adj. Chi-Square	1	0.2632	0.6080
Mantel-Haenszel Chi-Square	1	0.2968	0.5859
Phi Coefficient		0.0073	
Contingency Coefficient		0.0073	
Cramer's V		0.0073	

Fisher's Exact Test			
Cell (1,1) Frequency (F)	2206		
Left-sided Pr <= F	0.7179		
Right-sided Pr >= F	0.3040		
Table Probability (P)	0.0219		
Two-sided Pr <= P	0.5888		

Sample Size = 5599
Frequency Missing = 38

Frequency Col Pct	Table of agegroup_ed by gp_record		
	gp_record		
agegroup_ed	No	Yes	Total
Under 60 years old	3149 72.32	1086 84.65	4235
60 years old and older	1205 27.68	197 15.35	1402
Total	4354	1283	5637

Statistics for Table of agegroup_ed by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	80.5115	<.0001
Likelihood Ratio Chi-Square	1	86.9179	<.0001
Continuity Adj. Chi-Square	1	79.8535	<.0001
Mantel-Haenszel Chi-Square	1	80.4972	<.0001
Phi Coefficient		-0.1195	
Contingency Coefficient		0.1187	
Cramer's V		-0.1195	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3149
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5637

Frequency
Col Pct

Table of cob_ed by gp_record			
cob_ed(Country of birth)	gp_record		
	No	Yes	Total
Australia	3203 77.14	586 45.67	3789
Overseas	949 22.86	697 54.33	1646
Total	4152	1283	5435

Frequency Missing = 202

Statistics for Table of cob_ed by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	459.7318	<.0001
Likelihood Ratio Chi-Square	1	433.4514	<.0001
Continuity Adj. Chi-Square	1	458.2425	<.0001
Mantel-Haenszel Chi-Square	1	459.6472	<.0001
Phi Coefficient		0.2908	
Contingency Coefficient		0.2793	
Cramer's V		0.2908	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3203
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5435
Frequency Missing = 202

Frequency
Col Pct

Table of health_insurance by gp_record			
health_insurance(Have private health insurance)	gp_record		
	No	Yes	Total
No	1914 45.42	576 46.01	2490
Yes	2300 54.58	676 53.99	2976
Total	4214	1252	5466

Frequency Missing = 171

Statistics for Table of health_insurance by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	0.1338	0.7145
Likelihood Ratio Chi-Square	1	0.1338	0.7146
Continuity Adj. Chi-Square	1	0.1112	0.7388
Mantel-Haenszel Chi-Square	1	0.1338	0.7145
Phi Coefficient		-0.0049	
Contingency Coefficient		0.0049	
Cramer's V		-0.0049	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	1914
Left-sided Pr <= F	0.3692
Right-sided Pr >= F	0.6549
Table Probability (P)	0.0241
Two-sided Pr <= P	0.7223

Sample Size = 5466
Frequency Missing = 171

Frequency
Col Pct

Table of smoker_ED by gp_record			
smoker_ED	gp_record		
	No	Yes	Total
No	4231 97.18	1191 92.83	5422
Yes, smoker	123 2.82	92 7.17	215
Total	4354	1283	5637

Statistics for Table of smoker_ED by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	51.0136	<.0001
Likelihood Ratio Chi-Square	1	44.2731	<.0001

Statistic	DF	Value	Prob
Continuity Adj. Chi-Square	1	49.8359	<.0001
Mantel-Haenszel Chi-Square	1	51.0045	<.0001
Phi Coefficient		0.0951	
Contingency Coefficient		0.0947	
Cramer's V		0.0951	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	4231
Left-sided Pr <= P	1.0000
Right-sided Pr >= P	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5637

Frequency Col Pct	Table of risky_alcohol_ED by gp_record		
	gp_record		Total
risky_alcohol_ED	No	Yes	
No	4224 97.01	1178 91.82	5402
Yes, drinker	130 2.99	105 8.18	235
Total	4354	1283	5637

Statistics for Table of risky_alcohol_ED by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	67.0261	<.0001
Likelihood Ratio Chi-Square	1	57.8864	<.0001
Continuity Adj. Chi-Square	1	65.7313	<.0001
Mantel-Haenszel Chi-Square	1	67.0142	<.0001
Phi Coefficient		0.1090	
Contingency Coefficient		0.1084	
Cramer's V		0.1090	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	4224
Left-sided Pr <= P	1.0000
Right-sided Pr >= P	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 5637

Frequency Col Pct	Table of obesity_ED by gp_record		
	gp_record		Total
obesity_ED	No	Yes	
No	4174 95.87	1196 93.22	5370
Yes, obese	180 4.13	87 6.78	267
Total	4354	1283	5637

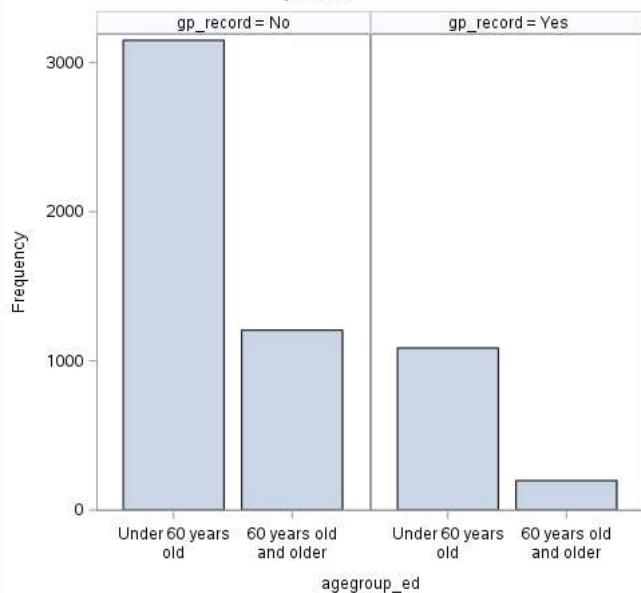
Statistics for Table of obesity_ED by gp_record

Statistic	DF	Value	Prob
Chi-Square	1	15.3864	<.0001
Likelihood Ratio Chi-Square	1	14.1930	0.0002
Continuity Adj. Chi-Square	1	14.8054	0.0001
Mantel-Haenszel Chi-Square	1	15.3837	<.0001
Phi Coefficient		0.0522	
Contingency Coefficient		0.0522	
Cramer's V		0.0522	

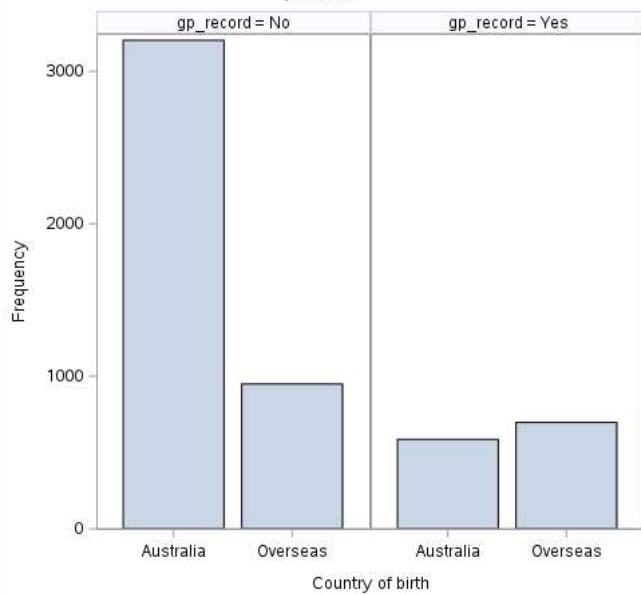
Fisher's Exact Test	
Cell (1,1) Frequency (F)	4174
Left-sided Pr <= P	0.9999
Right-sided Pr >= P	0.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	0.0002

Sample Size = 5637

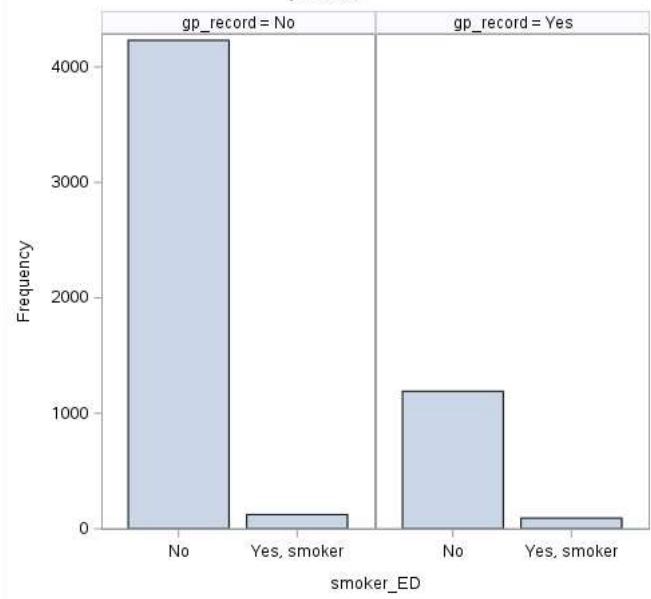
Health-related characteristics report of the highest 25% of ED visit patients



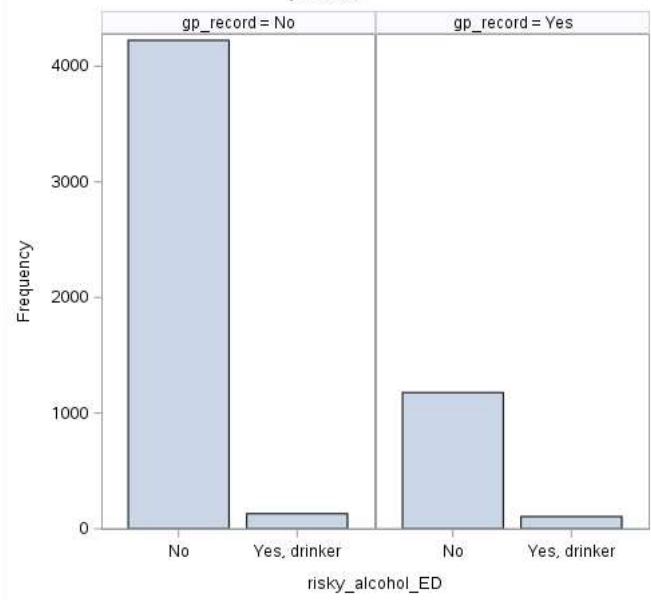
Health-related characteristics report of the highest 25% of ED visit patients



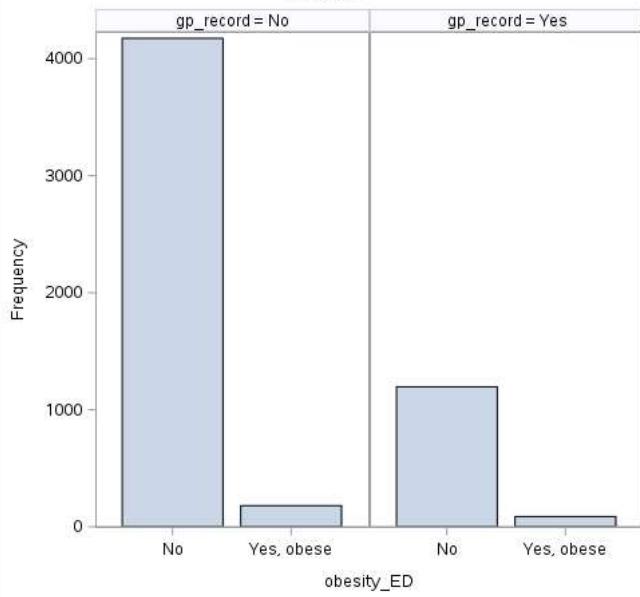
Health-related characteristics report of the highest 25% of ED visit patients



Health-related characteristics report of the highest 25% of ED visit patients



Health-related characteristics report of the highest 25% of ED visit patients



Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Sex of patient=.

Frequency

Table of smoker_ED by Smoke_current_GP		
smoker_ED	Smoke_current_GP(Being a current smoker)	
	No	Total
No	1	1
Total	1	1

Frequency Missing = 37

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Sex of patient=Male

Frequency

Table of smoker_ED by Smoke_current_GP			
smoker_ED	Smoke_current_GP(Being a current smoker)		
	No	Yes	Total
No	459	117	576
Yes, smoker	21	30	51
Total	480	147	627

Frequency Missing = 2223

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity			
Statistic	Estimate	Standard Error	95% Confidence Limits
Sensitivity	0.9563	0.0093	0.9380 0.9745
Specificity	0.2041	0.0332	0.1389 0.2692
Positive Predictive Value	0.7969	0.0168	0.7640 0.8297
Negative Predictive Value	0.5882	0.0689	0.4532 0.7233

Sample Size = 627

Frequency Missing = 2223

WARNING: 78% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Sex of patient=Female

Frequency

Table of smoker_ED by Smoke_current_GP			
smoker_ED	Smoke_current_GP(Being a current smoker)		
	No	Yes	Total
No	512	78	590
Yes, smoker	21	20	41

Table of smoker_ED by Smoke_current_GP				
smoker_ED	Smoke_current_GP(Being a current smoker)			Total
	No	Yes		
Total	533	98		631
Frequency Missing = 2118				

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9606	0.0084	0.9441	0.9771
Specificity	0.2041	0.0407	0.1243	0.2839
Positive Predictive Value	0.8678	0.0139	0.8405	0.8951
Negative Predictive Value	0.4878	0.0781	0.3348	0.6408

Sample Size = 631
Frequency Missing = 2118

WARNING: 77% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

agegroup_ed=Under 60 years old

Frequency	Table of smoker_ED by Smoke_current_GP				
smoker_ED	Smoke_current_GP(Being a current smoker)			Total	
	No	Yes			
No	813	173		986	
Yes, smoker	34	46		80	
Total	847	219		1066	
Frequency Missing = 3169					

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9599	0.0067	0.9466	0.9731
Specificity	0.2100	0.0275	0.1561	0.2640
Positive Predictive Value	0.8245	0.0121	0.8008	0.8483
Negative Predictive Value	0.5750	0.0553	0.4667	0.6833

Sample Size = 1066
Frequency Missing = 3169

WARNING: 75% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

agegroup_ed=60 years old and older

Frequency	Table of smoker_ED by Smoke_current_GP				
smoker_ED	Smoke_current_GP(Being a current smoker)			Total	
	No	Yes			
No	159	22		181	
Yes, smoker	8	4		12	
Total	167	26		193	
Frequency Missing = 1209					

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9521	0.0165	0.9197	0.9845
Specificity	0.1538	0.0708	0.0152	0.2925
Positive Predictive Value	0.8785	0.0243	0.8308	0.9261
Negative Predictive Value	0.3333	0.1361	0.0666	0.6001

Sample Size = 193
Frequency Missing = 1209

WARNING: 86% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Country of birth=.

For smoker_ED * Smoke_current_GP
 all data are missing because
 all levels of variable Smoke_current_GP are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Country of birth=Australia

Frequency

smoker_ED	Table of smoker_ED by Smoke_current_GP		
	Smoke_current_GP(Being a current smoker)		
	No	Yes	Total
No	503	45	548
Yes, smoker	18	12	30
Total	521	57	578

Frequency Missing = 3211

Statistics for Table of smoker_ED by Smoke_current_GP

Statistic	Sensitivity and Specificity			
	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9655	0.0080	0.9498	0.9811
Specificity	0.2105	0.0540	0.1047	0.3164
Positive Predictive Value	0.9179	0.0117	0.8949	0.9409
Negative Predictive Value	0.4000	0.0894	0.2247	0.5753

Sample Size = 578

Frequency Missing = 3211

WARNING: 85% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Country of birth=Overseas

Frequency

smoker_ED	Table of smoker_ED by Smoke_current_GP		
	Smoke_current_GP(Being a current smoker)		
	No	Yes	Total
No	469	150	619
Yes, smoker	24	38	62
Total	493	188	681

Frequency Missing = 965

Statistics for Table of smoker_ED by Smoke_current_GP

Statistic	Sensitivity and Specificity			
	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9513	0.0097	0.9323	0.9703
Specificity	0.2021	0.0293	0.1447	0.2595
Positive Predictive Value	0.7577	0.0172	0.7239	0.7914
Negative Predictive Value	0.6129	0.0619	0.4917	0.7341

Sample Size = 681

Frequency Missing = 965

WARNING: 59% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Have private health insurance=.

Frequency

smoker_ED	Table of smoker_ED by Smoke_current_GP		
	Smoke_current_GP(Being a current smoker)		
	No	Yes	Total
No	22	5	27
Yes, smoker	1	3	4
Total	23	8	31

Frequency Missing = 140

Statistics for Table of smoker_ED by Smoke_current_GP

Statistic	Sensitivity and Specificity			
	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9565	0.0425	0.8732	1.0000
Specificity	0.3750	0.1712	0.0395	0.7105
Positive Predictive Value	0.8148	0.0748	0.6683	0.9613
Negative Predictive Value	0.7500	0.2165	0.3257	1.0000

Sample Size = 31
Frequency Missing = 140

WARNING: 82% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Have private health insurance=No

Frequency

		Table of smoker_ED by Smoke_current_GP		
		Smoke_current_GP(Being a current smoker)		
smoker_ED		No	Yes	Total
		442	93	535
Yes, smoker		18	14	32
Total		460	107	567
Frequency Missing = 1923				

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9609	0.0090	0.9431	0.9786
Specificity	0.1308	0.0326	0.0669	0.1947
Positive Predictive Value	0.8262	0.0164	0.7941	0.8583
Negative Predictive Value	0.4375	0.0877	0.2656	0.6094

Sample Size = 567
Frequency Missing = 1923

WARNING: 77% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Have private health insurance=Yes

Frequency

		Table of smoker_ED by Smoke_current_GP		
		Smoke_current_GP(Being a current smoker)		
smoker_ED		No	Yes	Total
		508	97	605
Yes, smoker		23	33	56
Total		531	130	661
Frequency Missing = 2315				

Statistics for Table of smoker_ED by Smoke_current_GP

Sensitivity and Specificity				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Sensitivity	0.9567	0.0088	0.9394	0.9740
Specificity	0.2538	0.0382	0.1790	0.3287
Positive Predictive Value	0.8397	0.0149	0.8104	0.8689
Negative Predictive Value	0.5893	0.0657	0.4604	0.7181

Sample Size = 661
Frequency Missing = 2315

WARNING: 78% of the data are missing.

Health-related characteristics report of the highest 25% of ED visit patients

gp_smoker
757

Health-related characteristics report of the highest 25% of ED visit patients

ed_smoker
215

Health-related characteristics report of the highest 25% of ED visit patients

gped_smoker
922

Health-related characteristics report of the highest 25% of ED visit patients

The CONTENTS Procedure

Data Set Name	WORK.COHORT	Observations	922
Member Type	DATA	Variables	27
Engine	V9	Indexes	0
Created	21/02/2024 20:48:17	Observation Length	224
Last Modified	21/02/2024 20:48:17	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	YES
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	584
Obs in First Data Page	558
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5A680001DAF4_odaws02-apse1-2.oda.sas.com/SAS_workC01E0001DAF4_odaws02-apse1-2.oda.sas.com/cohort.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	546207
Access Permission	rwr--r--
Owner Name	u62298616
File Size	384KB
File Size (bytes)	393216

Variables in Creation Order				
#	Variable	Type	Len	Format
1	ID	Num	8	Unique patientID
2	GP_last	Num	8	DDMMYY10. Date of most recent GP visit
3	healthcare_card	Num	8	YNF. Have a health care card
4	drinks_day	Num	8	Number of alcohol drinks per day
5	height	Num	8	Body height (m)
6	weight	Num	8	Body weight (kg)
7	adverse_reaction	Num	8	YNF. Had any reaction to aby medication
8	syst_bp	Num	8	Systolic blood pressure (mmHg)
9	diast_bp	Num	8	Diastolic blood pressure (mmHg)
10	reason	Char	16	Reason for the most recent GP visit
11	Smoke_current_GP	Num	8	YNF. Being a current smoker
12	Risky_alcohol_GP	Num	8	YNF. Have two or more alcohol drinks per day
13	BMI_GP	Num	8	YNF. BMI score
14	Obese_GP	Num	8	YNF. Being obese (BMI>=30)
15	HighBP_GP	Num	8	YNF. Have high blood pressure (>=135/85mmHg)
16	Agegroup_GP	Num	8	AGEGROUP.
17	ed_admission	Num	8	DDMMYY10. Date of ED attendance
18	ed_separation	Num	8	DDMMYY10. Date of ED separation
19	sex_ed	Num	8	SEXF. Sex of patient
20	cob_ed	Num	8	COBF. Country of birth
21	interpreter	Num	8	YNF. Interpreter required
22	health_insurance	Num	8	YNF. Have private health insurance
23	triage_category	Num	8	TRIAGEF. Urgency of presentation
24	separation_mode	Num	8	SEPMODEF. Status at separation from ED
25	smoker_ED	Num	8	SMOKER.
26	risky_alcohol_ED	Num	8	DRINKER.
27	obesity_ED	Num	8	OBSE.

Sort Information	
Sortedby	ID
Validated	YES
Character Set	ASCII
Sort Option	NODUPKEY

Health-related characteristics report of the highest 25% of ED visit patients

Obs	ID	GP_last	healthcare_card	drinks_day	height	weight	adverse_reaction	syst_bp	diast_bp	reason	Smoke_current_GP	Risky_alcohol_GP	BMI_GP	Obese_GP	HighBP_GP	Agegroup_GP
1	15	01/07/2014	No	.	1.75	83.6	No	117	69	HEADACHE	Yes	.	27	No	No	Under 60 years old
2	16	16/02/2014	No	0	1.74	90.3	No	128	86	TINNITUS	Yes	No	30	No	Yes	Under 60 years old
3	18	05/03/2014	No	0	1.73	95.7	No	131	88	HEADACHE	Yes	No	32	Yes	Yes	Under 60 years old
4	23	13/05/2014	Yes	3	1.72	100.6	No	140	92	HEADACHE	Yes	Yes	34	Yes	Yes	Under 60 years old
5	32	18/02/2014	No	3	1.75	84.4	No	115	73	HEADACHE	Yes	Yes	28	No	No	Under 60 years old
6	33	03/08/2014	Yes	0	1.64	99.2	No	104	65	HEADACHE	No	No	37	Yes	No	60 years old and older
7	43	25/06/2014	No	1	1.81	120.6	No	112	66	HEADACHE	Yes	No	37	Yes	No	Under 60 years old

Obs	ID	GP_last	healthcare_card	drinks_day	height	weight	adverse_reaction	syst_bp	diast_bp	reason	Smoke_current_GP	Risky_alcohol_GP	BMI_GP	Obese_GP	HighBP_GP	Agegroup_GP
8	47	29/08/2014	Yes	0	1.60	65.6	No	114	84	NAUSEA	Yes	No	26	No	No	Under 60 years old
9	57	31/08/2014	No	2	1.77	57.3	Yes	118	70	HEADACHE	Yes	No	18	No	No	Under 60 years old
10	58	30/04/2014	No	0	1.68	54.6	No	95	65	HEADACHE	Yes	No	19	No	No	Under 60 years old

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Unique person ID				
ID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15	1	0.32	1	0.32
41	1	0.32	2	0.63
100	1	0.32	3	0.95
101	1	0.32	4	1.27
128	1	0.32	5	1.59
154	1	0.32	6	1.90
178	1	0.32	7	2.22
215	1	0.32	8	2.54
354	1	0.32	9	2.86
384	1	0.32	10	3.17
450	1	0.32	11	3.49
476	1	0.32	12	3.81
520	1	0.32	13	4.13
593	1	0.32	14	4.44
632	1	0.32	15	4.76
672	1	0.32	16	5.08
693	1	0.32	17	5.40
705	2	0.63	19	6.03
719	1	0.32	20	6.35
746	1	0.32	21	6.67
917	1	0.32	22	6.98
944	1	0.32	23	7.30
1088	1	0.32	24	7.62
1100	1	0.32	25	7.94
1117	1	0.32	26	8.25
1136	1	0.32	27	8.57
1172	1	0.32	28	8.89
1228	3	0.95	31	9.84
1291	1	0.32	32	10.16
1330	1	0.32	33	10.48
1336	1	0.32	34	10.79
1341	1	0.32	35	11.11
1342	1	0.32	36	11.43
1425	1	0.32	37	11.75
1484	1	0.32	38	12.06
1512	1	0.32	39	12.38
1541	1	0.32	40	12.70
1545	1	0.32	41	13.02
1638	1	0.32	42	13.33
1707	1	0.32	43	13.65
1722	1	0.32	44	13.97
1767	1	0.32	45	14.29
1802	1	0.32	46	14.60
1872	1	0.32	47	14.92
1873	1	0.32	48	15.24
1929	1	0.32	49	15.56
1935	1	0.32	50	15.87
1942	1	0.32	51	16.19
1945	1	0.32	52	16.51
1958	2	0.63	54	17.14
1975	1	0.32	55	17.46
2029	1	0.32	56	17.78
2061	1	0.32	57	18.10
2095	1	0.32	58	18.41
2105	1	0.32	59	18.73
2110	4	1.27	63	20.00
2128	2	0.63	65	20.63
2187	1	0.32	66	20.95
2194	1	0.32	67	21.27
2218	1	0.32	68	21.59
2340	1	0.32	69	21.90
2373	6	1.90	75	23.81
2379	12	3.81	87	27.62
2451	1	0.32	88	27.94
2513	2	0.63	90	28.57
2528	1	0.32	91	28.89
2538	4	1.27	95	30.16

Unique person ID				
ID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2540	1	0.32	96	30.48
2555	1	0.32	97	30.79
2571	1	0.32	98	31.11
2597	1	0.32	99	31.43
2711	1	0.32	100	31.75
2739	1	0.32	101	32.06
2756	1	0.32	102	32.38
2776	6	1.90	108	34.29
2799	1	0.32	109	34.60
2856	1	0.32	110	34.92
2949	1	0.32	111	35.24
2951	1	0.32	112	35.56
2956	2	0.63	114	36.19
2963	2	0.63	116	36.83
2966	5	1.59	121	38.41
2974	1	0.32	122	38.73
2981	2	0.63	124	39.37
3149	3	0.95	127	40.32
3158	1	0.32	128	40.63
3190	2	0.63	130	41.27
3257	3	0.95	133	42.22
3290	1	0.32	134	42.54
3305	1	0.32	135	42.86
3330	3	0.95	138	43.81
3346	1	0.32	139	44.13
3410	2	0.63	141	44.76
3472	1	0.32	142	45.08
3481	1	0.32	143	45.40
3497	1	0.32	144	45.71
3514	1	0.32	145	46.03
3521	1	0.32	146	46.35
3538	3	0.95	149	47.30
3555	4	1.27	153	48.57
3586	1	0.32	154	48.89
3595	1	0.32	155	49.21
3691	7	2.22	162	51.43
3798	2	0.63	164	52.06
3904	1	0.32	165	52.38
3913	1	0.32	166	52.70
3955	3	0.95	169	53.65
3956	4	1.27	173	54.92
3992	3	0.95	176	55.87
4101	1	0.32	177	56.19
4209	1	0.32	178	56.51
4258	2	0.63	180	57.14
4307	1	0.32	181	57.46
4332	2	0.63	183	58.10
4344	1	0.32	184	58.41
4384	1	0.32	185	58.73
4400	1	0.32	186	59.05
4421	7	2.22	193	61.27
4427	1	0.32	194	61.59
4451	1	0.32	195	61.90
4470	1	0.32	196	62.22
4517	1	0.32	197	62.54
4552	1	0.32	198	62.86
4556	1	0.32	199	63.17
4619	1	0.32	200	63.49
4697	1	0.32	201	63.81
4717	1	0.32	202	64.13
4721	1	0.32	203	64.44
4757	1	0.32	204	64.76
4763	8	2.54	212	67.30
4772	2	0.63	214	67.94
4773	1	0.32	215	68.25
4815	1	0.32	216	68.57
4846	1	0.32	217	68.89
4891	1	0.32	218	69.21
4947	1	0.32	219	69.52
4956	1	0.32	220	69.84
4966	1	0.32	221	70.16
4971	2	0.63	223	70.79
4973	1	0.32	224	71.11
4974	1	0.32	225	71.43
5027	1	0.32	226	71.75
5044	1	0.32	227	72.06
5085	1	0.32	228	72.38
5093	2	0.63	230	73.02
5100	1	0.32	231	73.33

Unique person ID				
ID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
5159	1	0.32	232	73.65
5171	1	0.32	233	73.97
5235	1	0.32	234	74.29
5268	1	0.32	235	74.60
5363	1	0.32	236	74.92
5371	1	0.32	237	75.24
5395	1	0.32	238	75.56
5411	1	0.32	239	75.87
5429	1	0.32	240	76.19
5498	1	0.32	241	76.51
5538	1	0.32	242	76.83
5564	1	0.32	243	77.14
5566	1	0.32	244	77.46
5573	1	0.32	245	77.78
5615	1	0.32	246	78.10
5623	1	0.32	247	78.41
5625	1	0.32	248	78.73
5646	1	0.32	249	79.05
5725	1	0.32	250	79.37
5735	1	0.32	251	79.68
5783	1	0.32	252	80.00
5786	1	0.32	253	80.32
10101	1	0.32	254	80.63
10179	1	0.32	255	80.95
10241	2	0.63	257	81.59
10272	1	0.32	258	81.90
10462	1	0.32	259	82.22
10538	1	0.32	260	82.54
10577	1	0.32	261	82.86
10890	1	0.32	262	83.17
10898	1	0.32	263	83.49
10980	1	0.32	264	83.81
11000	1	0.32	265	84.13
11053	1	0.32	266	84.44
11101	1	0.32	267	84.76
11205	1	0.32	268	85.08
11234	1	0.32	269	85.40
11247	1	0.32	270	85.71
11614	1	0.32	271	86.03
11637	1	0.32	272	86.35
11735	2	0.63	274	86.98
12097	1	0.32	275	87.30
12300	1	0.32	276	87.62
12400	1	0.32	277	87.94
12650	1	0.32	278	88.25
12904	2	0.63	280	88.89
12915	2	0.63	282	89.52
13047	1	0.32	283	89.84
13158	1	0.32	284	90.16
13362	2	0.63	286	90.79
13410	2	0.63	288	91.43
13465	4	1.27	292	92.70
13480	2	0.63	294	93.33
13799	1	0.32	295	93.65
13875	2	0.63	297	94.29
14107	1	0.32	298	94.60
14120	1	0.32	299	94.92
14241	1	0.32	300	95.24
14362	1	0.32	301	95.56
14391	1	0.32	302	95.87
14407	1	0.32	303	96.19
14437	2	0.63	305	96.83
14495	1	0.32	306	97.14
14566	2	0.63	308	97.78
14700	2	0.63	310	98.41
14743	1	0.32	311	98.73
15008	1	0.32	312	99.05
15120	1	0.32	313	99.37
15496	1	0.32	314	99.68
15580	1	0.32	315	100.00

Anatomical Therapeutic Chemical code				
ATC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N06AX12	1	0.32	1	0.32
N07BA01	79	25.08	80	25.40

Anatomical Therapeutic Chemical code				
ATC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N07BA03	235	74.60	315	100.00

Health-related characteristics report of the highest 25% of ED visit patients

The FREQ Procedure

Anatomical Therapeutic Chemical code				
ATC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N07BA01	61	26.07	61	26.07
N07BA03	173	73.93	234	100.00
Frequency Missing = 770				

