CPH576D- Assignment 2

Dominic LaRoche

September 17, 2014

1 Introduction

For this section we need to calculate the eGFR score for cases by both race and gender for each time point in the data: 1 month, 6 months, 1 year, and 3 years. This code include modified code from assignment 2 since many of the objectives were similar. Below is the code calculating the eGFR scores followed by the summary tables and cross-tabulations requested.

2 SAS Code

```
/*create dummy variables for race and female*/
data gfr;
set class.test_sample_3b;
if (r_gender = 'U' or r_gender = '') then r_gender = .; /*set U or blank to missing in gender*/
if r_gender = . then fem = .;
if r_gender = 'F' then fem = 1;
if r_gender = 'M' then fem = 0;
if r_race = 'AfrAmeri' then blk = 1;
  else blk = 0;
run;

/*use arrays to loop through each CR measurement
and calculate eGFR*/
data egfr;
set gfr;
array cr{4} CR_1M--CR_3Y;
array eg{4} egfr_1M egfr_6M egfr_1Y egfr_3Y;
do i=1 to 4;
egfr_base = 186*(cr{i}**-1.154)*(r_age**-0.203);
if (fem = 1 and blk = 1) then eg{i} = egfr_base*1.21*0.742;
if (fem = 1 and blk = 0) then eg{i} = egfr_base*0.742;
if (fem = 0 and blk = 1) then eg{i} = egfr_base*1.21;
if (fem = 0 and blk = 0) then eg{i} = egfr_base;
end;
drop i egfr_base;
run;


/* ceate ckd stages- care must be taken since the range of egfr
is outside the values for ckd stage.
```

```
I will create a stage 0 which indicates healthy kidney function.
Also missing values are considered very
small numbers so I will have to be careful of those*/

data ckd;
set egfr;
if (egfr_1M > 100) then ckdstg = 0;
if (90 le egfr_1M le 100) then ckdstg = 1;
if (60 le egfr_1M lt 90) then ckdstg = 2;
if (30 le egfr_1M lt 60) then ckdstg = 3;
if (15 le egfr_1M lt 30) then ckdstg = 4;
if (0 le egfr_1M lt 15) then ckdstg = 5;
run;

/*Create dichotomous kidney function indicator
for males and females separately*/

data ckd;
set ckd;
array eg{4} egfr_1M -- egfr_3Y;
array kf{4} kfunc_1M kfunc_6M kfunc_1Y kfunc_3Y;
do i=1 to 4;

if (fem = 1) and (1 < eg{i} < 70) then kf{i} = 0;
if (fem = 1) and (70 < eg{i} >= 70) then kf{i} = 1;
if (fem = 0) and (1< eg{i} < 68) then kf{i} = 0;
if (fem = 0) and (68 < eg{i} >= 68) then kf{i} = 1;
end;
drop i;
run;

proc freq data=class.test_sample_3b;
tables r_age r_gender r_race;
run;

proc univariate data=egfr ;
var egfr_1M;
histogram;
run;

proc freq data=ckd;
tables ckdstg;
run;

proc freq data=ckd;
tables kfunc_1M*kfunc_3Y;
run;
```

3 Output

The FREQ Procedure

r_age

r_age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-1	1	0.33	1	0.33
2	1	0.33	2	0.65
3	1	0.33	3	0.98
20	1	0.33	4	1.30
23	1	0.33	5	1.63
24	1	0.33	6	1.95
26	2	0.65	8	2.61
27	6	1.95	14	4.56
28	4	1.30	18	5.86
29	5	1.63	23	7.49
30	1	0.33	24	7.82
31	5	1.63	29	9.45
32	5	1.63	34	11.07
33	6	1.95	40	13.03
34	10	3.26	50	16.29
35	12	3.91	62	20.20
36	9	2.93	71	23.13
37	4	1.30	75	24.43
38	9	2.93	84	27.36
39	16	5.21	100	32.57
40	11	3.58	111	36.16
41	9	2.93	120	39.09
42	14	4.56	134	43.65
43	11	3.58	145	47.23
44	11	3.58	156	50.81
45	9	2.93	165	53.75
46	13	4.23	178	57.98
47	11	3.58	189	61.56
48	10	3.26	199	64.82
49	7	2.28	206	67.10
50	9	2.93	215	70.03
51	6	1.95	221	71.99
52	9	2.93	230	74.92
53	10	3.26	240	78.18
54	4	1.30	244	79.48
55	7	2.28	251	81.76
56	6	1.95	257	83.71
57	5	1.63	262	85.34
58	3	0.98	265	86.32
59	6	1.95	271	88.27
60	5	1.63	276	89.90
61	3	0.98	279	90.88
62	2	0.65	281	91.53
63	9	2.93	290	94.46
64	4	1.30	294	95.77
65	4	1.30	298	97.07
67	4	1.30	302	98.37
68	2	0.65	304	99.02
70	1	0.33	305	99.35
72	1	0.33	306	99.67
74	1	0.33	307	100.00

r_gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	111	36.27	111	36.27
M	192	62.75	303	99.02
U	3	0.98	306	100.00

Frequency Missing = 1

r_race	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AfrAmeri	3	0.98	3	0.98
AmIndian	9	2.95	12	3.93
Caucasia	292	95.74	304	99.67
Other	1	0.33	305	100.00

Frequency Missing = 2

The UNIVARIATE Procedure
Variable: egfr_1M

Moments

N	290	Sum Weights	290
Mean	53.9208838	Sum Observations	15637.0563
Std Deviation	17.847589	Variance	318.536432
Skewness	1.08197961	Kurtosis	3.3437137
Uncorrected SS	935220.923	Corrected SS	92057.0287
Coeff Variation	33.0995854	Std Error Mean	1.04804651

Basic Statistical Measures

Location		Variability	
Mean	53.92088	Std Deviation	17.84759
Median	51.63445	Variance	318.53643
Mode	61.29646	Range	138.67177
		Interquartile Range	21.91894

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	51.44894	Pr > |t|	<.0001
Sign	M	145	Pr >= |M|	<.0001
Signed Rank	S	21097.5	Pr >= |S|	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	150.1119
99%	113.2275
95%	86.7046
90%	73.9412
75% Q3	63.8887
50% Median	51.6344
25% Q1	41.9698
10%	34.5962
5%	29.9688
1%	14.8842
0% Min	11.4402

Extreme Observations

------Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
11.4402	16	106.283	9
12.7044	13	111.108	96
14.8842	166	113.228	223
18.1951	52	114.391	230
19.5155	218	150.112	158

Missing Values

		-----Percent Of-----	
Missing Value	Count	All Obs	Missing Obs
.	17	5.54	100.00

The FREQ Procedure

ckdstg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	5	1.72	5	1.72
1	7	2.41	12	4.14
2	83	28.62	95	32.76
3	179	61.72	274	94.48
4	13	4.48	287	98.97
5	3	1.03	290	100.00

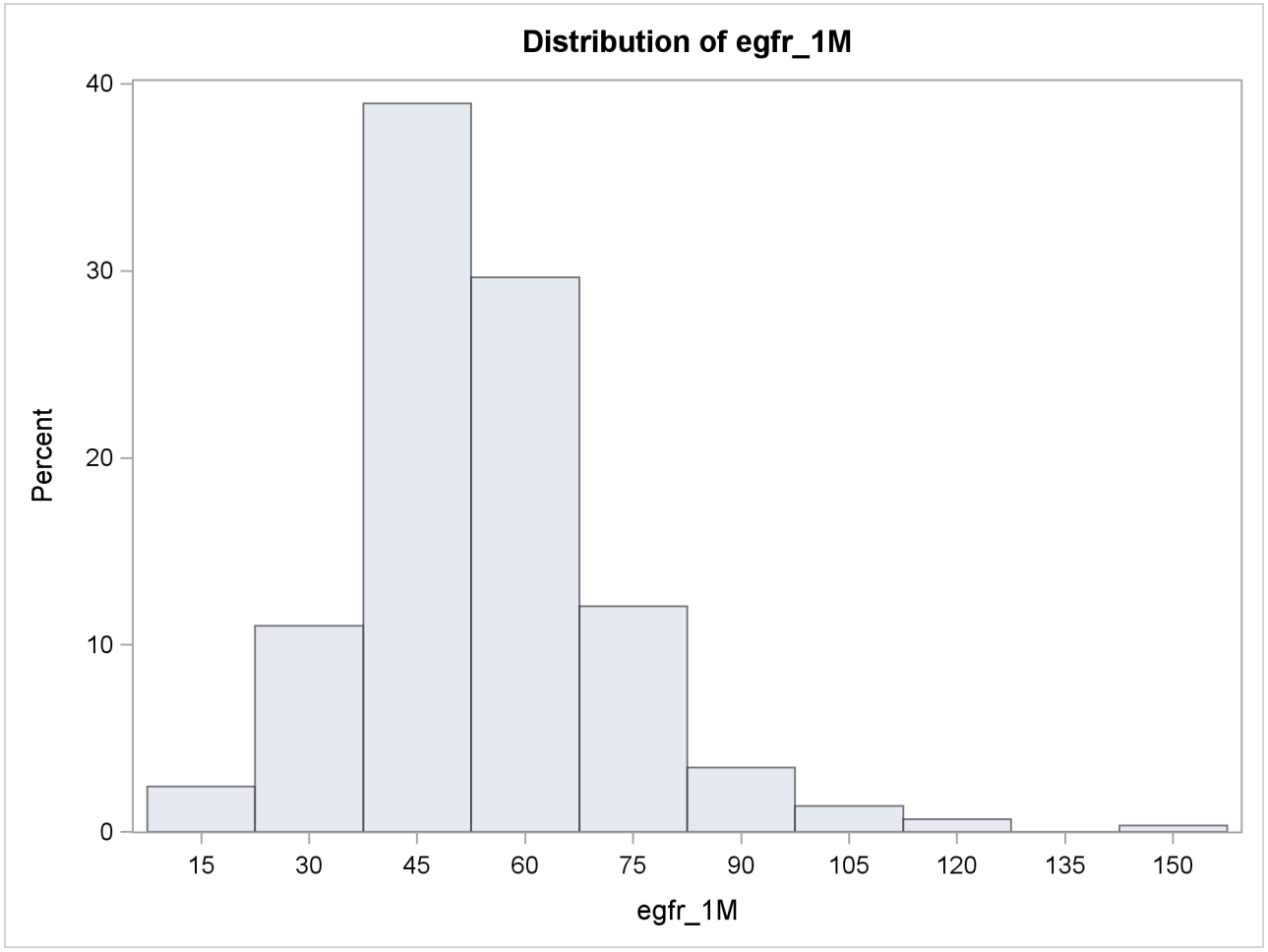
Frequency Missing = 17

The FREQ Procedure

Table of kfunc_1M by kfunc_3Y

kfunc_1M / kfunc_3Y	0	1	Total
0	201	19	220
	76.72	7.25	83.97
	91.36	8.64	
	89.73	50.00	
1	23	19	42
	8.78	7.25	16.03
	54.76	45.24	
	10.27	50.00	
Total	224	38	262
	85.50	14.50	100.00

Frequency Missing = 45


Distribution of egfr_1M
Percent
40
30
20
10
0
15
30
45
60
75
90
105
120
135
150
egfr_1M

4 Log Output

```
1    /*
2     This file is auto-generated by the statrep package.
3     Do not edit this file or your changes will be lost.
4     Edit the LaTeX file instead.
5
6     See the statrep package documentation and the file
7     statrep.cfg for information on these settings.
8     */
9
10
11   %include "LaRoche_Assignment3_CPH576D_SR_preamble.sas" /nosource;
1441  /* Remove all output files. */
1442  %hostdel
1443
1444  /* Start program with a null title. */
1445  title;
1446
1447  /*create dummy variables for race and female*/
1448  data gfr;
1449  set class.test_sample_3b;
1450  if (r_gender = 'U' or r_gender = '') then r_gender = .; /*set U or blank
1450! to missing in gender*/
1451  if r_gender = . then fem = .;
1452  if r_gender = 'F' then fem = 1;
1453  if r_gender = 'M' then fem = 0;
1454  if r_race = 'AfrAmeri' then blk = 1;
1455    else blk = 0;
1456  run;

NOTE: Numeric values have been converted to character
      values at the places given by: (Line):(Column).
      1450:54
NOTE: Character values have been converted to numeric
      values at the places given by: (Line):(Column).
      1451:4
NOTE: Invalid numeric data, r_gender='F' , at line 1451 column 4.
r_age=50 diabetes_type=2 group=N Cont CR_1M=1.30 CR_6M=1.20 CR_1Y=2.10
CR_3Y=1.70 r_race=AfrAmeri r_gender=F fem=1 blk=1 _ERROR_=1 _N_=1
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=48 diabetes_type=1 group=N Cont CR_1M=. CR_6M=. CR_1Y=. CR_3Y=.
r_race=AfrAmeri r_gender=M fem=0 blk=1 _ERROR_=1 _N_=2
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=48 diabetes_type=2 group=PAK Tx CR_1M=2.00 CR_6M=1.80 CR_1Y=2.10
CR_3Y=1.90 r_race=AfrAmeri r_gender=M fem=0 blk=1 _ERROR_=1 _N_=3
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=67 diabetes_type=2 group=N Cont CR_1M=1.30 CR_6M=1.30 CR_1Y=1.60
CR_3Y=1.70 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=4
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=47 diabetes_type=2 group=N Cont CR_1M=1.50 CR_6M=1.80 CR_1Y=1.40
CR_3Y=1.30 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=5
```

```
NOTE: Invalid numeric data, r_gender='F' , at line 1451 column 4.
r_age=44 diabetes_type=2 group=N Cont CR_1M=. CR_6M=. CR_1Y=. CR_3Y=.
r_race=AmIndian r_gender=F fem=1 blk=0 _ERROR_=1 _N_=6
NOTE: Invalid numeric data, r_gender='F' , at line 1451 column 4.
r_age=53 diabetes_type=1 group=N Cont CR_1M=0.90 CR_6M=0.80 CR_1Y=0.80
CR_3Y=0.80 r_race=AmIndian r_gender=F fem=1 blk=0 _ERROR_=1 _N_=7
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=54 diabetes_type=1 group=N Cont CR_1M=1.20 CR_6M=2.00 CR_1Y=1.80
CR_3Y=2.30 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=8
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=56 diabetes_type=2 group=PAK Tx CR_1M=0.80 CR_6M=0.90 CR_1Y=0.90
CR_3Y=0.90 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=9
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=38 diabetes_type=1 group=PAK Tx CR_1M=2.50 CR_6M=1.40 CR_1Y=1.60
CR_3Y=1.50 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=10
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=42 diabetes_type=1 group=PAK Tx CR_1M=1.30 CR_6M=1.50 CR_1Y=2.40
CR_3Y=1.80 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=11
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=54 diabetes_type=1 group=Y Cont CR_1M=1.60 CR_6M=1.40 CR_1Y=2.10
CR_3Y=2.30 r_race=AmIndian r_gender=M fem=0 blk=0 _ERROR_=1 _N_=12
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=52 diabetes_type=2 group=N Cont CR_1M=5.10 CR_6M=2.70 CR_1Y=2.50
CR_3Y=3.70 r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=13
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=63 diabetes_type=1 group=N Cont CR_1M=2.10 CR_6M=1.50 CR_1Y=2.10
CR_3Y=2.40 r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=14
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=42 diabetes_type=1 group=N Cont CR_1M=5.80 CR_6M=. CR_1Y=. CR_3Y=.
r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=16
NOTE: Invalid numeric data, r_gender='F' , at line 1451 column 4.
r_age=39 diabetes_type=1 group=N Cont CR_1M=1.20 CR_6M=1.10 CR_1Y=1.20
CR_3Y=1.10 r_race=Caucasia r_gender=F fem=1 blk=0 _ERROR_=1 _N_=17
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=45 diabetes_type=1 group=N Cont CR_1M=2.30 CR_6M=2.00 CR_1Y=. CR_3Y=2.10
r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=18
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=68 diabetes_type=2 group=N Cont CR_1M=2.40 CR_6M=2.60 CR_1Y=4.20
CR_3Y=1.80 r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=19
NOTE: Invalid numeric data, r_gender='M' , at line 1451 column 4.
r_age=55 diabetes_type=1 group=N Cont CR_1M=2.40 CR_6M=1.70 CR_1Y=2.20
CR_3Y=2.10 r_race=Caucasia r_gender=M fem=0 blk=0 _ERROR_=1 _N_=20
NOTE: Invalid numeric data, r_gender='F' , at line 1451 column 4.
WARNING: Limit set by ERRORS= option reached.  Further errors of this type will
         not be printed.
r_age=49 diabetes_type=2 group=N Cont CR_1M=1.80 CR_6M=1.60 CR_1Y=1.72 CR_3Y=.
r_race=Caucasia r_gender=F fem=1 blk=0 _ERROR_=1 _N_=21
NOTE: There were 307 observations read from the data set CLASS.TEST_SAMPLE_3B.
NOTE: The data set WORK.GFR has 307 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.05 seconds
```

```
      cpu time            0.04 seconds


1457
1458  /*use arrays to loop through each CR measurement
1459  and calculate eGFR*/
1460  data egfr;
1461  set gfr;
1462  array cr{4} CR_1M--CR_3Y;
1463  array eg{4} egfr_1M egfr_6M egfr_1Y egfr_3Y;
1464  do i=1 to 4;
1465  egfr_base = 186*(cr{i}**-1.154)*(r_age**-0.203);
1466  if (fem = 1 and blk = 1) then eg{i} = egfr_base*1.21*0.742;
1467  if (fem = 1 and blk = 0) then eg{i} = egfr_base*0.742;
1468  if (fem = 0 and blk = 1) then eg{i} = egfr_base*1.21;
1469  if (fem = 0 and blk = 0) then eg{i} = egfr_base;
1470  end;
1471  drop i egfr_base;
1472  run;

NOTE: Invalid argument(s) to the exponential operator "**" at line 1465 column
      39.
NOTE: Invalid argument(s) to the exponential operator "**" at line 1465 column
      39.
NOTE: Invalid argument(s) to the exponential operator "**" at line 1465 column
      39.
NOTE: Invalid argument(s) to the exponential operator "**" at line 1465 column
      39.
r_age=-1 diabetes_type=2 group=Y Cont CR_1M=1.50 CR_6M=1.70 CR_1Y=2.10
CR_3Y=1.70 r_race=Caucasia r_gender=M fem=0 blk=0 egfr_1M=. egfr_6M=. egfr_1Y=.
egfr_3Y=. i=5 egfr_base=. _ERROR_=1 _N_=243
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).
      102 at 1465:16   102 at 1465:23   4 at 1465:32     33 at 1467:48
      4 at 1468:48
NOTE: Mathematical operations could not be performed at the following places.
      The results of the operations have been set to missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      4 at 1465:39
NOTE: There were 307 observations read from the data set WORK.GFR.
NOTE: The data set WORK.EGFR has 307 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds


1473
1474
1475  /* ceate ckd stages- care must be taken since the range of egfr
1476  is outside the values for ckd stage.
```

```
1477  I will create a stage 0 which indicates healthy kidney function.
1478  Also missing values are considered very
1479  small numbers so I will have to be careful of those*/
1480
1481  data ckd;
1482  set egfr;
1483  if (egfr_1M > 100) then ckdstg = 0;
1484  if (90 le egfr_1M le 100) then ckdstg = 1;
1485  if (60 le egfr_1M lt 90) then ckdstg = 2;
1486  if (30 le egfr_1M lt 60) then ckdstg = 3;
1487  if (15 le egfr_1M lt 30) then ckdstg = 4;
1488  if (0 le egfr_1M lt 15) then ckdstg = 5;
1489  run;

NOTE: There were 307 observations read from the data set WORK.EGFR.
NOTE: The data set WORK.CKD has 307 observations and 16 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


1490
1491  /*Create dichotomous kidney function indicator
1492  for males and females separately*/
1493
1494  data ckd;
1495  set ckd;
1496  array eg{4} egfr_1M -- egfr_3Y;
1497  array kf{4} kfunc_1M kfunc_6M kfunc_1Y kfunc_3Y;
1498  do i=1 to 4;
1499
1500  if (fem = 1) and (1 < eg{i} < 70) then kf{i} = 0;
1501  if (fem = 1) and (70 < eg{i} >= 70) then kf{i} = 1;
1502  if (fem = 0) and (1< eg{i} < 68) then kf{i} = 0;
1503  if (fem = 0) and (68 < eg{i} >= 68) then kf{i} = 1;
1504  end;
1505  drop i;
1506  run;

NOTE: There were 307 observations read from the data set WORK.CKD.
NOTE: The data set WORK.CKD has 307 observations and 20 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


1507
1508  %output(gfr)
1509  proc freq data=class.test_sample_3b;
1510  tables r_age r_gender r_race;
1511  run;
```

```
NOTE: There were 307 observations read from the data set CLASS.TEST_SAMPLE_3B.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds


1512
1513  proc univariate data=egfr ;
1514  var egfr_1M;
1515  histogram;
1516  run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds


1517
1518  proc freq data=ckd;
1519  tables ckdstg;
1520  run;

NOTE: There were 307 observations read from the data set WORK.CKD.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds


1521
1522  proc freq data=ckd;
1523  tables kfunc_1M*kfunc_3Y;
1524  run;

NOTE: There were 307 observations read from the data set WORK.CKD.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds


1525
1526  %endoutput(gfr)
1527
1528  %write(gfra,store=gfr,type=listing)
NOTE: Processing document gfr.
Note: New page for \Freq#1\Table1#1\OneWayFreqs#1
Note: New page for \Univariate#1\egfr_1M#1\Moments#1
Note: New page for \Freq#2\Table1#1\OneWayFreqs#1
Note: New page for \Freq#3\Table1#1\CrossTabFreqs#1
Note: Writing Listing file  : lst/gfra.lst
Note: Writing Listing file  : lst/gfra1.lst
```

```
Note: Writing Listing file  : lst/gfra2.lst
Note: Writing Listing file  : lst/gfra3.lst
Note: Writing Listing file  : lst/gfra4.lst
Note: Writing Listing file  : lst/gfra5.lst
Note: Writing Listing file  : lst/gfra6.lst
Note: Writing Listing file  : lst/gfra7.lst
Note: Writing Listing file  : lst/gfra8.lst
Note: Writing Listing file  : lst/gfra9.lst
Note: Writing Listing file  : lst/gfra10.lst

Objects                                  Type     Status   Group

Freq.Table1.OneWayFreqs                  Table    Selected   1
Freq.Table2.OneWayFreqs                  Table    Selected   2
Freq.Table3.OneWayFreqs                  Table    Selected   3
Univariate.egfr_1M.Moments               Table    Selected   4
Univariate.egfr_1M.BasicMeasures         Table    Selected   5
Univariate.egfr_1M.TestsForLocation      Table    Selected   6
Univariate.egfr_1M.Quantiles             Table    Selected   7
Univariate.egfr_1M.ExtremeObs            Table    Selected   8
Univariate.egfr_1M.MissingValues         Table    Selected   9
Univariate.egfr_1M.Histogram.Histogram   Graph                .
Freq#2.Table1.OneWayFreqs                Table    Selected  10
Freq#3.Table1.CrossTabFreqs              Crosstab Selected  11
1529
1530  %write(gfra,store=gfr,type=graphic)
NOTE: Processing document gfr.
NOTE: Writing Graph file: png/gfra.png

Objects                                  Type     Status   Group

Freq.Table1.OneWayFreqs                  Table                .
Freq.Table2.OneWayFreqs                  Table                .
Freq.Table3.OneWayFreqs                  Table                .
Univariate.egfr_1M.Moments               Table                .
Univariate.egfr_1M.BasicMeasures         Table                .
Univariate.egfr_1M.TestsForLocation      Table                .
Univariate.egfr_1M.Quantiles             Table                .
Univariate.egfr_1M.ExtremeObs            Table                .
Univariate.egfr_1M.MissingValues         Table                .
Univariate.egfr_1M.Histogram.Histogram   Graph    Selected    1
Freq#2.Table1.OneWayFreqs                Table                .
Freq#3.Table1.CrossTabFreqs              Crosstab             .
```