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. The primary variable of interest (Creatinine) was mesaured at four time-points but for this assignment we will only use the first time-point. Below is the code calculating the eGFR scores followed by the summary tables and plots.

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;

  egfr_base = 186*(CR_1M**-1.154)*(r_age**-0.203);
if (fem = 1 and blk = 1) then egfr = egfr_base*1.21*0.742;
if (fem = 1 and blk = 0) then egfr = egfr_base*0.742;
if (fem = 0 and blk = 1) then egfr = egfr_base*1.21;
if (fem = 0 and blk = 0) then egfr = egfr_base;

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


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

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

3 Output

The FREQ Procedure

diabetes_type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	247	80.46	247	80.46
2	60	19.54	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

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)

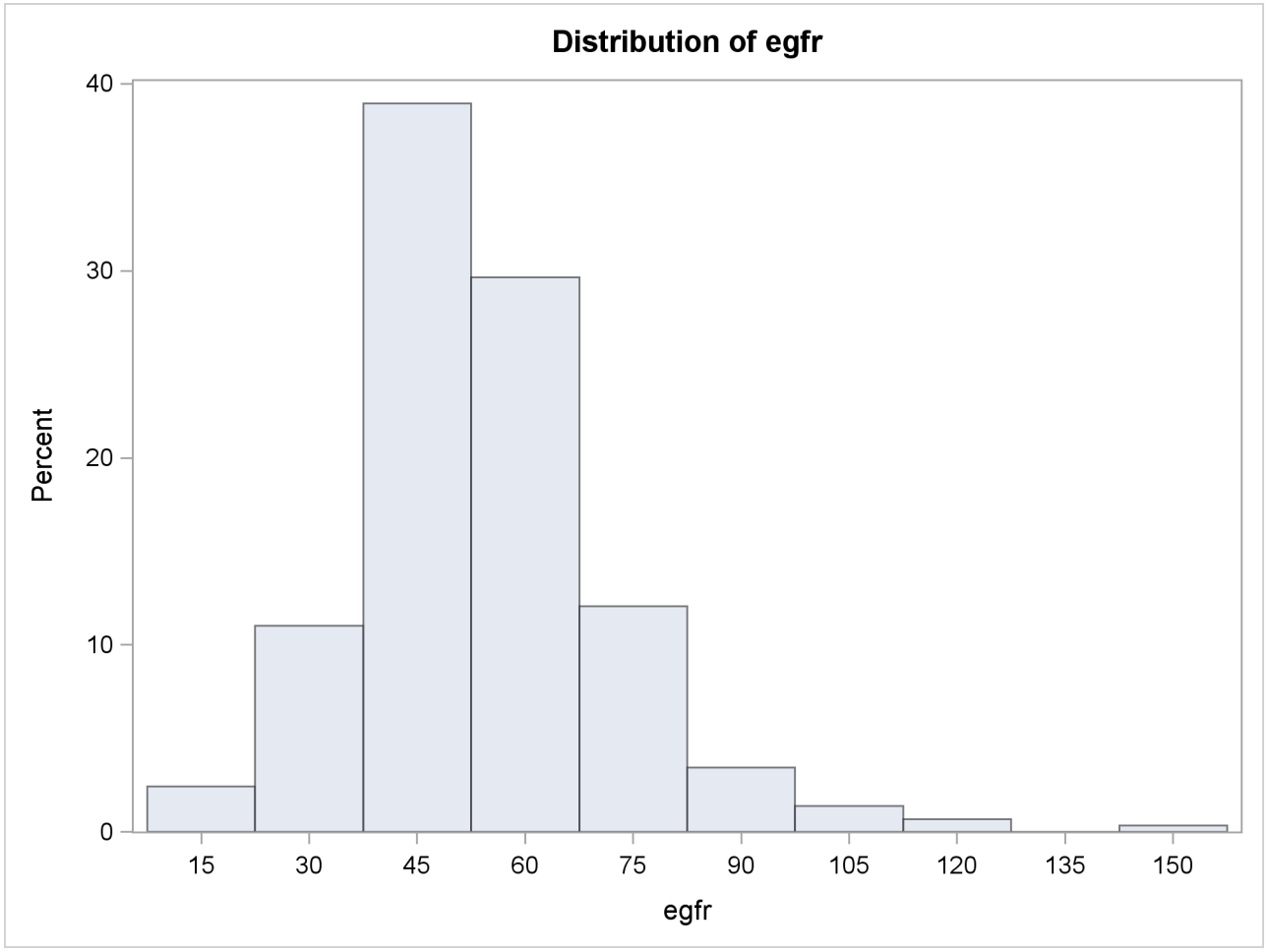
Quantile	Estimate
100% Max	150.1119
99%	113.2275
95%	86.7046
90%	73.9412
75% Q3	63.8887
50% Median	51.6344
25% Q1	41.9698
10%	34.5962
5%	29.9688
1%	14.8842
0% Min	11.4402

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


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

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_Assignment2_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           0.04 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
1463    egfr_base = 186*(CR_1M**-1.154)*(r_age**-0.203);
1464  if (fem = 1 and blk = 1) then egfr = egfr_base*1.21*0.742;
1465  if (fem = 1 and blk = 0) then egfr = egfr_base*0.742;
1466  if (fem = 0 and blk = 1) then egfr = egfr_base*1.21;
1467  if (fem = 0 and blk = 0) then egfr = egfr_base;
1468
1469  drop egfr_base;
1470  run;

NOTE: Invalid argument(s) to the exponential operator "**" at line 1463 column
      41.
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_base=. egfr=. _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).
      12 at 1463:18   12 at 1463:25   1 at 1463:34    5 at 1465:47
      1 at 1466:47
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).
      1 at 1463:41
NOTE: There were 307 observations read from the data set WORK.GFR.
NOTE: The data set WORK.EGFR has 307 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


1471
1472
1473  /* ceate ckd stages- care must be taken since the range of egfr
1474  is outside the values for ckd stage.
1475  I will create a stage 0 which indicates healthy kidney function.
1476  Also missing values are considered very
1477  small numbers so I will have to be careful of those*/
1478
1479  data ckd;
1480  set egfr;
1481  if (egfr > 100) then ckdstg = 0;
1482  if (90 le egfr le 100) then ckdstg = 1;
```

```
1483  if (60 le egfr lt 90) then ckdstg = 2;
1484  if (30 le egfr lt 60) then ckdstg = 3;
1485  if (15 le egfr lt 30) then ckdstg = 4;
1486  if (0 le egfr lt 15) then ckdstg = 5;
1487  run;

NOTE: There were 307 observations read from the data set WORK.EGFR.
NOTE: The data set WORK.CKD has 307 observations and 13 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


1488
1489
1490  %output(gfr)
1491  proc freq data=class.test_sample_3b;
1492  tables diabetes_type r_gender r_race;
1493  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            0.01 seconds


1494
1495  proc univariate data=egfr ;
1496  var egfr;
1497  histogram;
1498  run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.06 seconds
      cpu time            0.01 seconds


1499
1500
1501
1502  %endoutput(gfr)
1503
1504  %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#1\Moments#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

Objects                              Type  Status   Group

Freq.Table1.OneWayFreqs              Table Selected   1
Freq.Table2.OneWayFreqs              Table Selected   2
Freq.Table3.OneWayFreqs              Table Selected   3
Univariate.egfr.Moments              Table Selected   4
Univariate.egfr.BasicMeasures        Table Selected   5
Univariate.egfr.TestsForLocation     Table Selected   6
Univariate.egfr.Quantiles            Table Selected   7
Univariate.egfr.ExtremeObs           Table Selected   8
Univariate.egfr.MissingValues        Table Selected   9
Univariate.egfr.Histogram.Histogram  Graph            .
1505
1506  %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.Moments              Table            .
Univariate.egfr.BasicMeasures        Table            .
Univariate.egfr.TestsForLocation     Table            .
Univariate.egfr.Quantiles            Table            .
Univariate.egfr.ExtremeObs           Table            .
Univariate.egfr.MissingValues        Table            .
Univariate.egfr.Histogram.Histogram  Graph Selected   1
```