BIMI-6100 Elements in Biomedical Informatics
Fall 2023

David Crosslin

Tulane University School of Medicine
John W. Deming Department of Medicine
Division of Biomedical Informatics and Genomics New Orleans, LA

March 11, 2024

Contents

List of Figures

List of Tables

1 Hmisc `describe` function - rhc data

rhc[, -1]
62 Variables 5735 Observations

cat1

n	missing	distinct
5735	0	9

ARF (2490, 0.434), CHF (456, 0.080), Cirrhosis (224, 0.039), Colon Cancer (7, 0.001), Coma (436, 0.076), COPD (457, 0.080), Lung Cancer (39, 0.007), MOSF w/Malignancy (399, 0.070), MOSF w/Sepsis (1227, 0.214)

cat2

n	missing	distinct
1200	4535	6

Value	Cirrhosis	Colon Cancer	Coma	Lung Cancer
Frequency	38	2	90	15
Proportion	0.032	0.002	0.075	0.013

Value	MOSF w/Malignancy	MOSF w/Sepsis
Frequency	229	826
Proportion	0.191	0.688

ca

n	missing	distinct
5735	0	3

Value	Metastatic	No	Yes
Frequency	384	4379	972
Proportion	0.067	0.764	0.169

sadmdte

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	1401	1	11639	589.7	10831	10912	11164	11759	12097	12298	12363

lowest : 10754 10755 10756 10757 10758, highest: 12437 12438 12439 12440 12441

dschdte

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5734	1	1486	1	11660	589.2	10854	10939	11184	11777	12120	12319	12386

lowest : 10757 10758 10759 10763 10765, highest: 12492 12493 12505 12524 12560

dthdte

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
3722	2013	1592	1	11754	619.4	10881	10991	11267	11832	12208	12418	12533

lowest : 10757 10758 10759 10763 10766, highest: 12772 12773 12778 12780 12783

lstctdte

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	932	1	11781	602.6	10942	11038	11316	11868	12244	12440	12522

lowest : 10756 10758 10762 10766 10768, highest: 12634 12638 12640 12642 12644

death

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	2013	3722
Proportion	0.351	0.649

---

cardiohx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.436	1013	0.1766	0.2909

---

chfhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.439	1021	0.178	0.2927

---

dementhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.266	564	0.09834	0.1774

---

psychhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.188	386	0.06731	0.1256

---

chrpulhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.461	1089	0.1899	0.3077

---

renalhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.127	255	0.04446	0.08499

---

liverhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.195	401	0.06992	0.1301

---

gibledhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.094	185	0.03226	0.06245

---

malighx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.53	1316	0.2295	0.3537

---

immunhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.59	1543	0.269	0.3934

---

transhx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.306	662	0.1154	0.2042

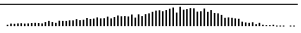
---

amihx

n	missing	distinct	Info	Sum	Mean	Gmd
5735	0	2	0.101	200	0.03487	0.06733

---

age



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	5036	1	61.38	18.86	30.02	36.97	50.15	64.05	73.93	80.88	85.00

lowest : 18.042 18.119 18.146 18.163 18.19 , highest: 97.84 98.089 100.246 100.849 101.848

---

sex

n	missing	distinct
5735	0	2

Value	Female	Male
Frequency	2543	3192
Proportion	0.443	0.557

---

edu

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	42	0.977	11.68	3.344	7	8	10	12	13	16	17

```
lowest :  0  1  2  3  4, highest: 24 25 26 27 30
```

surv2md1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	1522	1	0.5925	0.2177	0.2080	0.3100	0.4709	0.6280	0.7430	0.8165	0.8470

```
lowest : 0           0.000999928 0.00199986  0.002       0.00299978
highest: 0.938965    0.939       0.94        0.941       0.962
```

das2d3pc

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	1023	0.999	20.5	5.964	13.50	14.75	16.06	19.75	23.43	29.14	29.15

```
lowest : 11   12   12.5 13   13.5, highest: 30.5 31   31.5 32   33
```

t3d30

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	29	0.699	23.61	9.454	3	5	16	30	30	30	30

```
lowest :  2  3  4  5  6, highest: 26 27 28 29 30
```

dth30

n	missing	distinct
5735	0	2

```
Value          No   Yes
Frequency    3817  1918
Proportion 0.666 0.334
```

aps1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	123	1	54.67	22.45	24	30	41	54	67	81	90

```
lowest :   3   4   6   7   8, highest: 127 129 130 135 147
```

scoma1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	11	0.839	21	29.49	0	0	0	0	41	61	100

```
Value          0     9    26    37    41    44    55    61    89    94   100
Frequency   3105   497   383   259   214   466   132   175    62    81   361
Proportion 0.541 0.087 0.067 0.045 0.037 0.081 0.023 0.031 0.011 0.014 0.063

For the frequency table, variable is rounded to the nearest 1
```

meanbp1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	178	1	78.52	41.61	38	42	50	63	115	133	145

```
lowest :   0  10  13  15  17, highest: 210 212 222 242 259
```

wblc1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	520	1	15.65	11.3	1.700	4.300	8.398	14.100	20.049	27.500	34.000

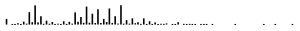
```
lowest : 0         0.0499954 0.0999908 0.109985  0.119995
highest: 100.391   112.188   171       178.594   192
```

hrt1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	189	1	115.2	45.1	45	58	97	124	141	160	171

```
lowest :   0   8  10  16  20, highest: 230 240 245 248 250
```

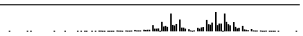
resp1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	72	0.997	28.09	15.87	8.0	8.4	14.0	30.0	38.0	44.0	50.0

```
lowest : 0       2       3       4       5       , highest: 72      80      90      94.2969 100
```

temp1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	118	0.999	37.62	1.984	34.90	35.40	36.09	38.09	39.00	39.59	40.00

```
lowest : 27      28      28.2969 29.5    30.1992, highest: 41.6953 41.7969 41.8984 42      43
```

pafi1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	1342	0.999	222.3	127.7	68.00	86.66	133.31	202.50	316.62	360.00	423.44

```
lowest : 11.5996 12      19      19.0469 20      , highest: 823.75  828.5   861.875 869.375 937.5
```

alb1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	57	0.916	3.093	0.7435	1.8	2.0	2.6	3.5	3.5	3.5	3.9

```
lowest : 0.299988 0.399963 0.5      0.699951 0.799927, highest: 5.7998   6.2998   6.59961  11       29
```

hema1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	450	1	31.87	9.214	21.00	23.00	26.10	30.00	36.30	44.59	47.03

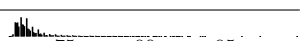
```
lowest : 2       4.5     7.69922 8.09961 9       , highest: 64.5    65.0938 65.5    65.8906 66.1875
```

bili1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	266	0.954	2.267	2.721	0.4000	0.5000	0.7999	1.0098	1.3999	4.0996	8.8984

```
lowest : 0.0999908 0.199982  0.299988  0.399963  0.5
highest: 50.1953   51.5      54.2969   56.1953   58.1953
```

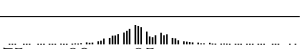
crea1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	148	0.999	2.133	1.786	0.600	0.700	1.000	1.500	2.400	4.300	6.199

```
lowest : 0.0999908 0.199982  0.299988  0.399963  0.5
highest: 17        17.0977   20.2969   21.8984   25.0977
```

sod1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	73	0.998	136.8	8.352	125	128	132	136	142	146	150

```
lowest : 101 102 103 105 106, highest: 172 173 174 177 178
```

pot1

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	81	0.998	4.067	1.09	2.8	3.0	3.4	3.8	4.6	5.5	6.0

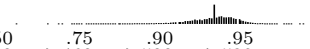
```
lowest : 1.09985 1.19995 1.3999  1.69995 1.7998 , highest: 9       9.29883 9.39844 10.3984 11.8984
```

paco21

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	266	0.998	38.75	12.98	23	26	31	37	42	53	63

```
lowest : 1       2       6       8       9       , highest: 133     148     151     155.281 156
```

ph1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	96	0.999	7.388	0.1162	7.189	7.250	7.340	7.400	7.460	7.500	7.529

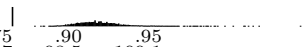
```
lowest : 6.5791  6.70996 6.75    6.76953 6.80957, highest: 7.69922 7.70996 7.71973 7.75    7.76953
```

swang1

n	missing	distinct
5735	0	2

Value	No RHC	RHC
Frequency	3551	2184
Proportion	0.619	0.381

wtkilo1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	922	0.999	67.83	30.6	0.0	38.5	56.3	70.0	83.7	98.5	109.1

```
lowest : 0     19.5  24.1  24.3  26.2 , highest: 200   200.8 204.5 209.2 244
```

dnr1

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5081	654
Proportion	0.886	0.114

ninsclas

n	missing	distinct
5735	0	6

Value	Medicaid	Medicare	Medicare & Medicaid	No insurance
Frequency	647	1458	374	322
Proportion	0.113	0.254	0.065	0.056

Value	Private	Private & Medicare
Frequency	1698	1236
Proportion	0.296	0.216

resp

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	3622	2113
Proportion	0.632	0.368

card

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	3804	1931
Proportion	0.663	0.337

---

neuro

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5042	693
Proportion	0.879	0.121

---

gastr

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	4793	942
Proportion	0.836	0.164

---

renal

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5440	295
Proportion	0.949	0.051

---

meta

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5470	265
Proportion	0.954	0.046

---

hema

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5381	354
Proportion	0.938	0.062

---

seps

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	4704	1031
Proportion	0.82	0.18

---

trauma

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5683	52
Proportion	0.991	0.009

---

ortho

n	missing	distinct
5735	0	2

Value	No	Yes
Frequency	5728	7
Proportion	0.999	0.001

---

adld3p

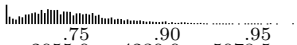
n	missing	distinct	Info	Mean	Gmd
1439	4296	8	0.832	1.182	1.699

Value	0.00	0.98	1.96	2.94	3.99	4.97	5.95	7.00
Frequency	778	296	130	43	49	64	53	26
Proportion	0.541	0.206	0.090	0.030	0.034	0.044	0.037	0.018

For the frequency table, variable is rounded to the nearest 0.07

---

urin1



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2707	3028	1440	1	2192	1655	120.0	473.8	1110.0	1927.0	2955.0	4230.0	5078.5

lowest : 0 2.11 3 3.52 3.69, highest: 8340 8370 8392 8580 9000

---

race

n	missing	distinct
5735	0	3

Value	black	other	white
Frequency	920	355	4460
Proportion	0.160	0.062	0.778

---

income

n	missing	distinct
5735	0	4

Value	> $50k	$11-$25k	$25-$50k	Under $11k
Frequency	451	1165	893	3226
Proportion	0.079	0.203	0.156	0.563

---

ptid

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
5735	0	5735	1	5134	3432	516.7	997.4	2561.5	5131.0	7689.0	9233.6	9786.9

lowest : 5 7 9 10 11, highest: 10270 10272 10273 10277 10278

Table 1: Descriptive Statistics by sex

	N	Female	Male	Combined	Test Statistic
		N = 2543	N = 3192	N = 5735	
death : Yes	5735	64% (1624)	66% (2098)	65% (3722)	χ^2_1=2.16, P=0.141[1]
race : black	5735	18% (465)	14% (455)	16% (920)	χ^2_2=17.2, P<0.001[1]
other		6% (157)	6% (198)	6% (355)	
white		76% (1921)	80% (2539)	78% (4460)	
age	5735	50.4 64.8 74.9	49.9 63.4 73.0	50.1 64.0 73.9	$F_{1,5733}$=9.63, P=0.002[2]
bili1	5735	0.70 1.01 1.10	0.90 1.01 1.60	0.80 1.01 1.40	$F_{1,5733}$=54.1, P<0.001[2]
crea1	5735	0.8 1.3 2.3	1.1 1.6 2.5	1.0 1.5 2.4	$F_{1,5733}$=101, P<0.001[2]

$_a$ b $_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. N is the number of non–missing values. Numbers after percents are frequencies. Tests used: [1]Pearson test; [2]Wilcoxon test

2 Hmisc `summary` function - rhc data

3 ols function output

Table 2: Effects Response : wblc1

	Low	High	Δ	Effect	S.E.	Lower 0.95	Upper 0.95
cardiohx	0	1	1	-0.934	0.412	-1.7420	-0.127
race — black:white	3	1		0.203	0.436	-0.6507	1.057
race — other:white	3	2		-2.077	0.654	-3.3590	-0.796
sex — Female:Male	2	1		0.521	0.316	-0.0998	1.141
income — > $50k:Under $11k	4	1		-1.059	0.600	-2.2349	0.117
income — $11—$25k:Under $11k	4	2		0.435	0.408	-0.3644	1.234
income — $25—$50k:Under $11k	4	3		-1.155	0.451	-2.0397	-0.271

Table 3: Analysis of Variance for `wblc1`

	d.f.	Partial SS	MS	F	P
race	2	1515	758	5.40	0.0045
sex	1	379	379	2.71	0.1000
income	3	1726	575	4.10	0.0064
cardiohx	1	721	721	5.14	0.0234
REGRESSION	7	4413	630	4.50	<0.0001
ERROR	5727	803152	140		

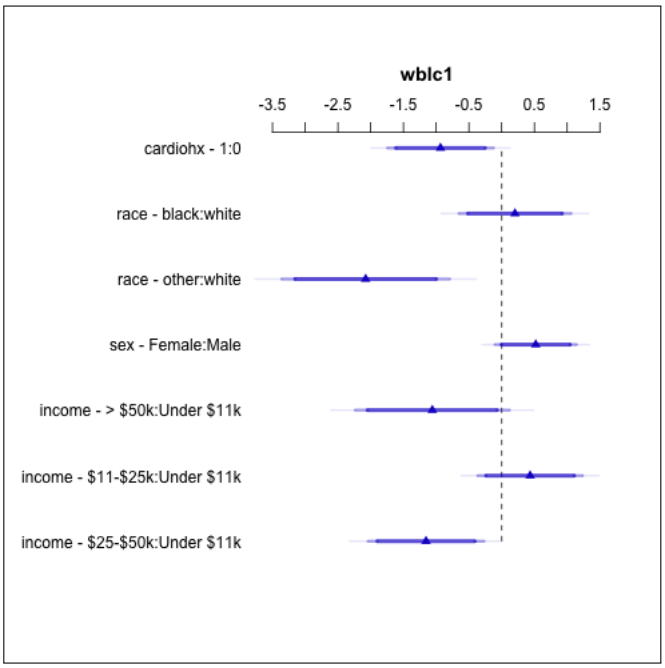


Figure 1: plot(summary(m))

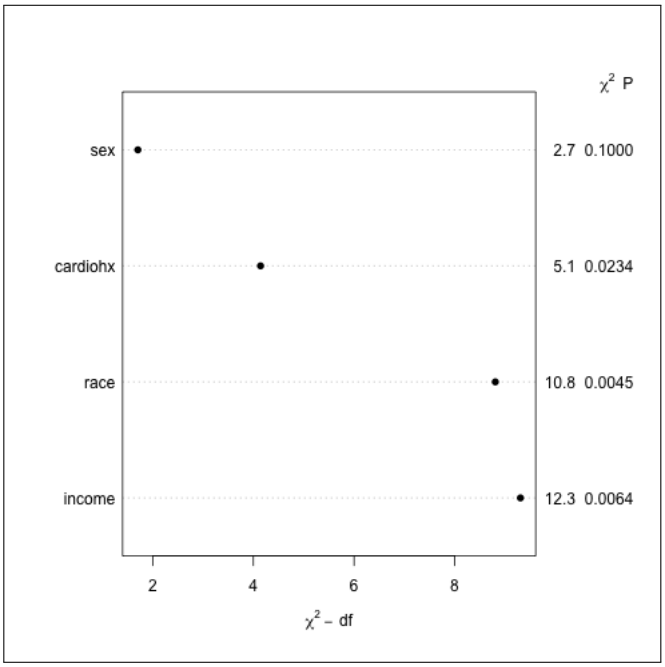


Figure 2: plot(anova(m))