

STAT 506 - Problem Set 1

Author: Zhen Qin, Uniqname: qinzhen

Notes: scripts and well formatted markdown files can be downloaded on [my github repo](#)/.

Problem1

a. Converting between long and wide data formats

The corresponding file and websites are **ProblemSet1_1(a).do** and <https://stats.idre.ucla.edu/stata/modules/> and <https://stats.idre.ucla.edu/stat/stata/modules/kids> , <https://jbhender.github.io/Stats506/dadmomlong.dta> , <https://stats.idre.ucla.edu/stat/stata/modules/dadmomw>.

b. Stata example 1

The corresponding file and websites are **ProblemSet1_1(b).do** and <https://stats.idre.ucla.edu/other/dae/> .

c. Stata example 2 The corresponding file and websites are **ProblemSet1_1(c).do** and <https://stats.idre.ucla.edu/other/dae/> .

Problem2

The corresponding file and websites are **ProblemSet1_2.do** and <http://www.eia.gov/consumption/residential/data/2009/> .

a.

`reportable~n` represents state and `rate` represents proportion of wood shingle roofs.

```
. list reportable~n rate if rooftype==2
```

	report~n	rate
67.	19	.0144062
81.	4	.0244547
84.	8	.0284647
86.	20	.0299494
87.	13	.0319221
88.	18	.0324584
89.	9	.0332125
92.	6	.0343415
96.	3	.0395206
97.	10	.039543
98.	1	.0398771
101.	11	.0450165
102.	2	.0478129
103.	24	.0511356
104.	12	.0521929
107.	14	.0551071
108.	15	.0573164
111.	7	.0592537
112.	5	.059617
113.	27	.0596553
117.	17	.0679545
122.	21	.0774461
127.	26	.0840876
130.	22	.0911869
131.	23	.092559
135.	25	.0981669
150.	16	.1246495

The max value is .1246495 and report~n is 16, the min value is .0144062 and reportable~n is 19. According to the codebook, the answer 1 is North Carolina, South Carolina, the answer 2 is Tennessee.

b.

A part of the table is shown below, which is the proportion of each roof type for all houses constructed in each decade. `rate` represents the proportion of each roof type for all houses constructed in each decade.

```
. list
```

	rooftype	decades	nweight	ss	rate
1.	4	1	459060.4	1.96e+07	.0233793
2.	2	1	988783.7	1.96e+07	.0503573

3.	1	1	195408.4	1.96e+07	.0099519
4.	-2	1	2723822	1.96e+07	.1387202
5.	5	1	9124287	1.96e+07	.4646863
6.	3	1	1160610	1.96e+07	.0591081
7.	8	1	400711.5	1.96e+07	.0204076
8.	6	1	4516175	1.96e+07	.230002
9.	7	1	66513.79	1.96e+07	.0033874
10.	5	2	7507134	1.35e+07	.5542037
11.	7	2	102747.1	1.35e+07	.0075852
12.	6	2	2625692	1.35e+07	.1938381
13.	8	2	165768	1.35e+07	.0122376
14.	-2	2	1279565	1.35e+07	.0944621
15.	1	2	129341.6	1.35e+07	.0095485
16.	2	2	984588.5	1.35e+07	.0726859
17.	3	2	556720	1.35e+07	.0410991
18.	4	2	194246.4	1.35e+07	.01434
19.	1	3	216477.3	1.33e+07	.0163117
20.	3	3	672852.3	1.33e+07	.0506997
21.	8	3	144047.2	1.33e+07	.010854
22.	4	3	143400.5	1.33e+07	.0108053
23.	-2	3	2565420	1.33e+07	.1933055
24.	2	3	880897.6	1.33e+07	.066376
25.	5	3	6308142	1.33e+07	.4753213
26.	7	3	74757.9	1.33e+07	.005633
27.	6	3	2265328	1.33e+07	.1706935
28.	3	4	1851955	1.83e+07	.1012532
29.	4	4	107987.8	1.83e+07	.0059041
30.	1	4	449815	1.83e+07	.024593
31.	5	4	7943300	1.83e+07	.4342896
32.	7	4	182489.9	1.83e+07	.0099774
33.	6	4	2506653	1.83e+07	.137048
34.	-2	4	4112612	1.83e+07	.2248517
35.	8	4	241359.4	1.83e+07	.013196
36.	2	4	894156.9	1.83e+07	.0488869
37.	-2	5	3494409	1.70e+07	.2059078
38.	5	5	7023901	1.70e+07	.413883
39.	7	5	194046.4	1.70e+07	.0114342
40.	4	5	161064.5	1.70e+07	.0094907
41.	1	5	675754.7	1.70e+07	.0398188
42.	8	5	88979.02	1.70e+07	.0052431
43.	6	5	2344584	1.70e+07	.1381544
44.	3	5	1857067	1.70e+07	.1094275
45.	2	5	1130937	1.70e+07	.0666404
46.	6	6	2259867	1.64e+07	.138217
47.	3	6	1600812	1.64e+07	.0979082

```
48 | -2 6 2499773 1 646407 15289 |
```

For full table, please execute the script.

The following table includes `relrate` (rate in the picture) representing relative rise between 1950 and 2000.

```
. list rooftype rate
```

	rooftype	rate
1.	4	1.616363
2.	8	1.797576
3.	6	2.657586
4.	5	4.003606
5.	2	4.791831
6.	-2	5.122133
7.	3	5.634457
8.	1	12.22678
9.	7	12.25835

Use this subset by decades to calculate, Concrete Tiles roof saw the largest relative rise.

Problem3

The corresponding file and websites are **ProblemSet1_3.do** , **AUX_D.DTA**, **DEMO_D.DTA** and <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Examination&CycleBeginYear=2005> and <https://wwwn.cdc.gov/Nchs/Nhanes/Search/DataPage.aspx?Component=Demographics&CycleBeginYear=2005> .

a.

By using command option, I can drop unnecessary cases.

b.

Above all, I replace AUXU* with log(AUXU*), because when decibel increases by 10db, volume increases by 10 times.

Compare each frequency separately and only need to use the first test at each frequency. Here is just an example.

```
. summarize AUXU1K1R
```

Variable	Obs	Mean	Std. Dev.	Min	Max
AUXU1K1R	2,746	2.926309	.7468581	0	6.801283

Use t test to examine it.

```
. ttest AUXU1K1R, by(group)
```

Two-sample t test with equal variances

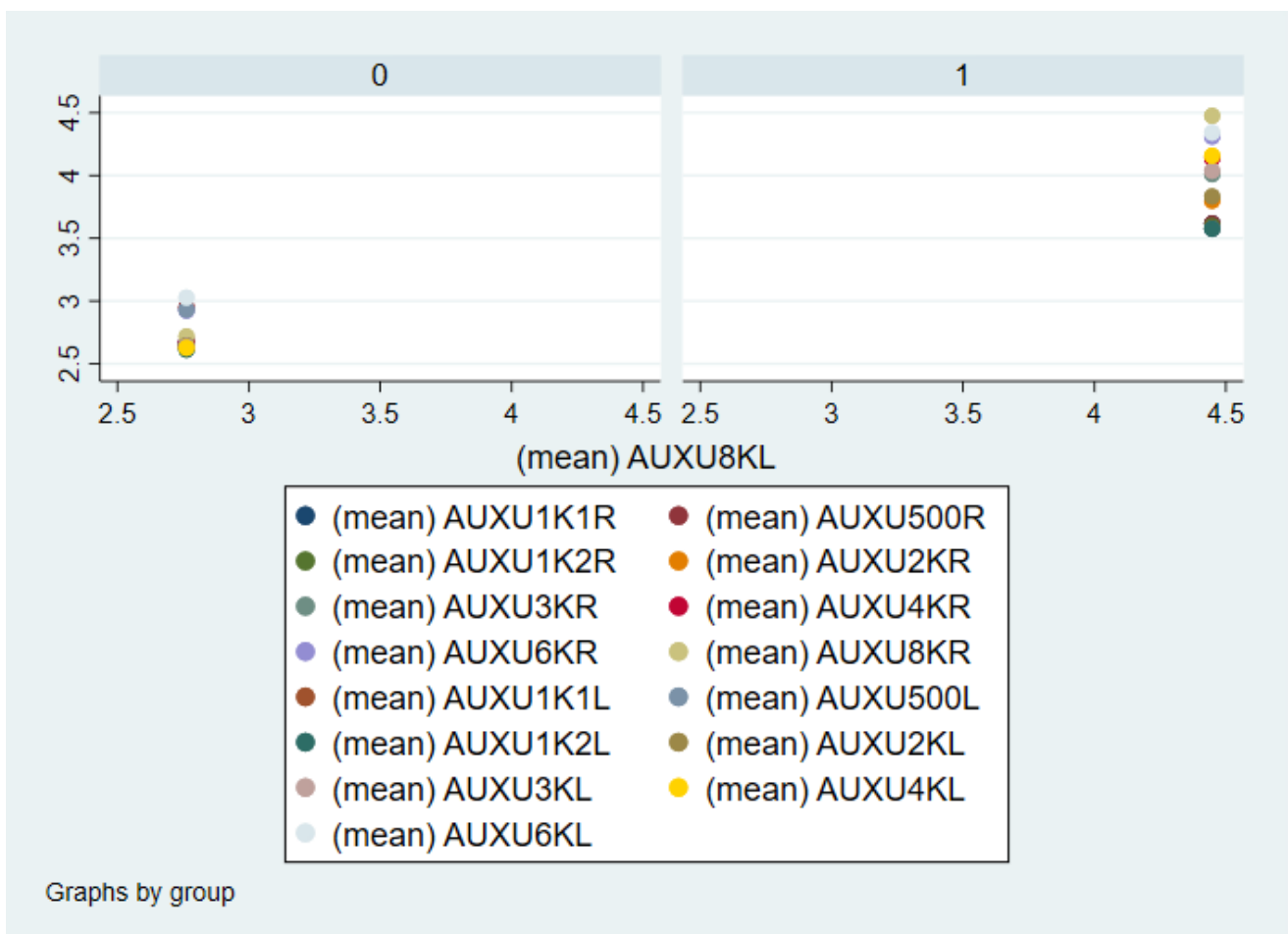
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	2,016	2.675796	.0144154	.6472498	2.647525	2.704066
1	730	3.618138	.0199162	.5381068	3.579038	3.657238
combined	2,746	2.926309	.0142524	.7468581	2.898363	2.954256
diff		-.942342	.0267872		-.9948671	-.889817

diff = mean(0) - mean(1) t = -35.1789
 Ho: diff = 0 degrees of freedom = 2744

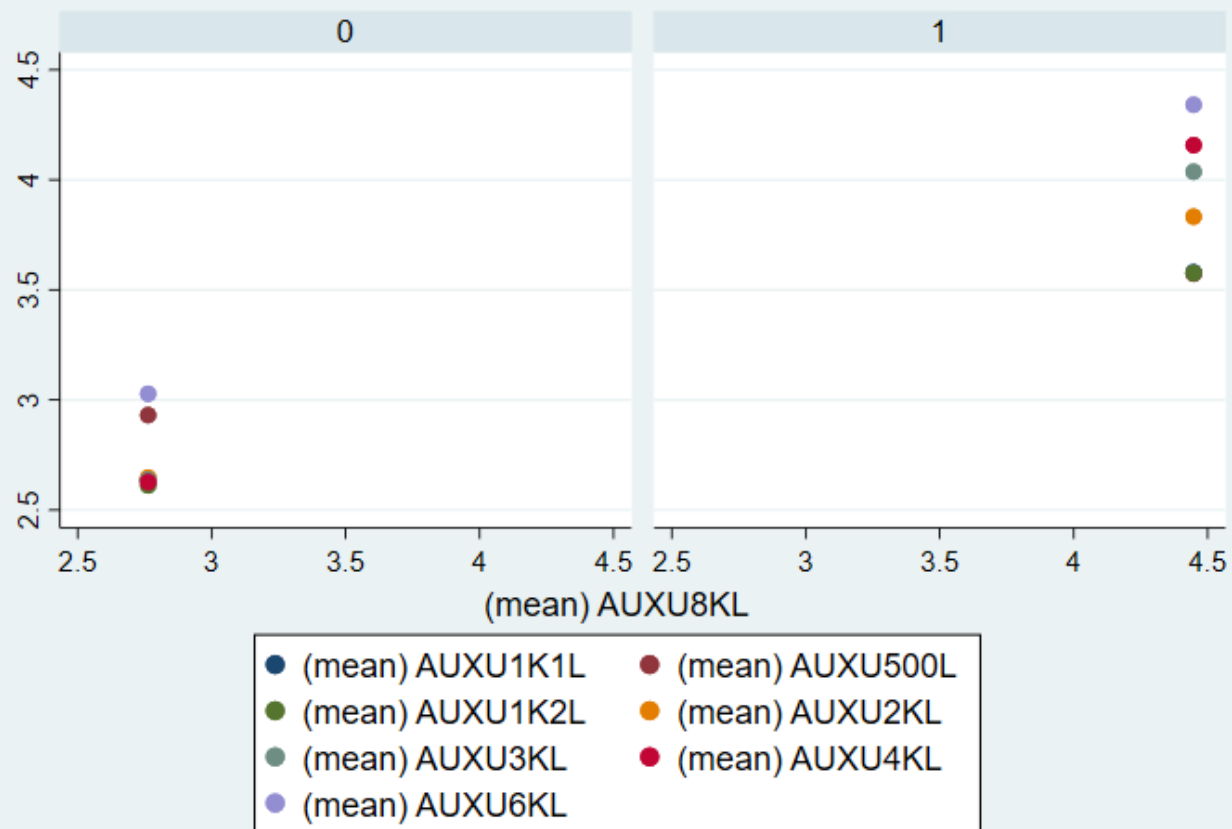
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Based on data, hearing loss due to age is common at each frequency because the two-sample t tests reject the null hypothesis: diff=0. Besides, neither ear is more prone to hearing loss.

For left and right ears:

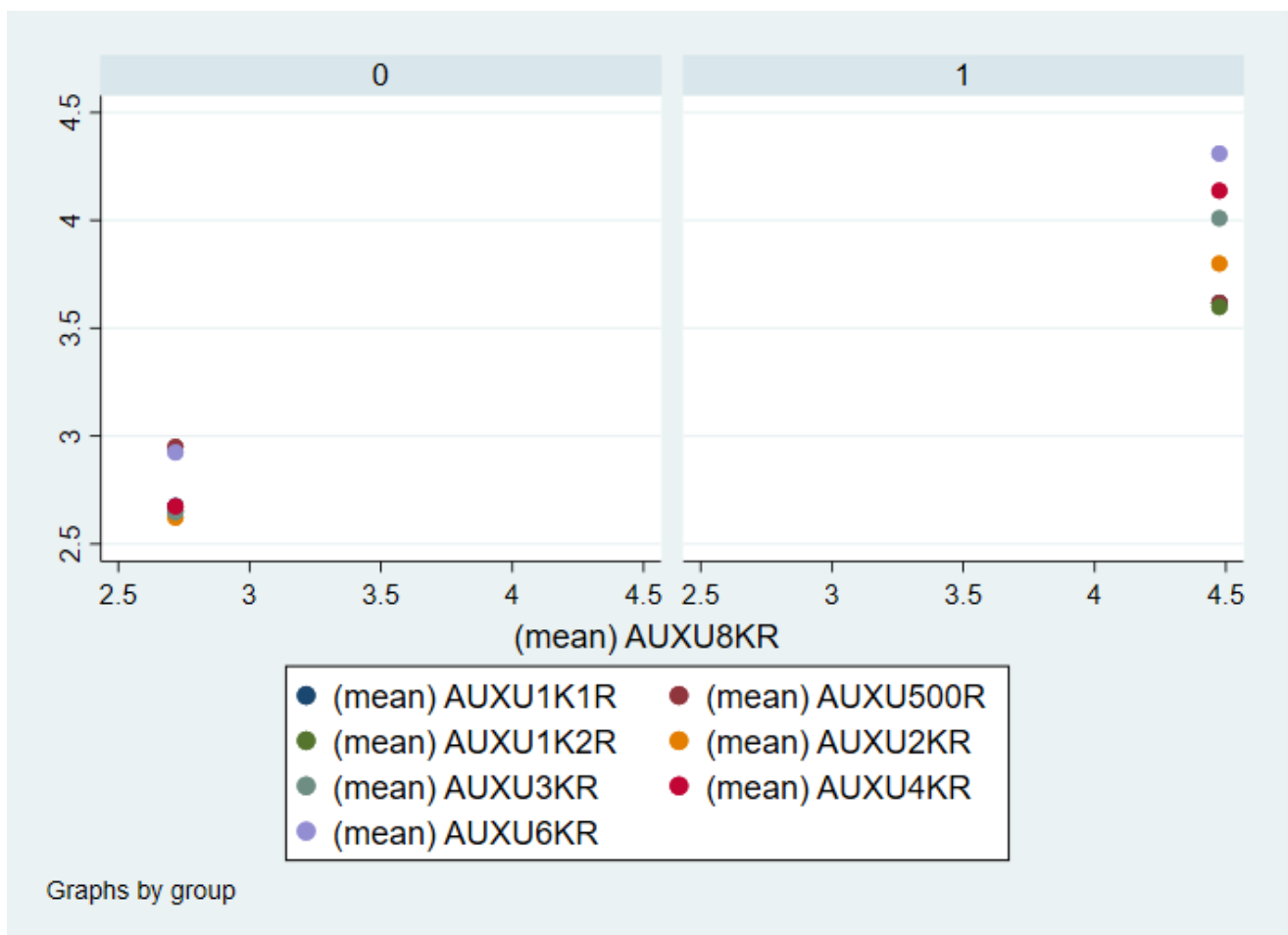


For the left ear:



Graphs by group

For the right ear:

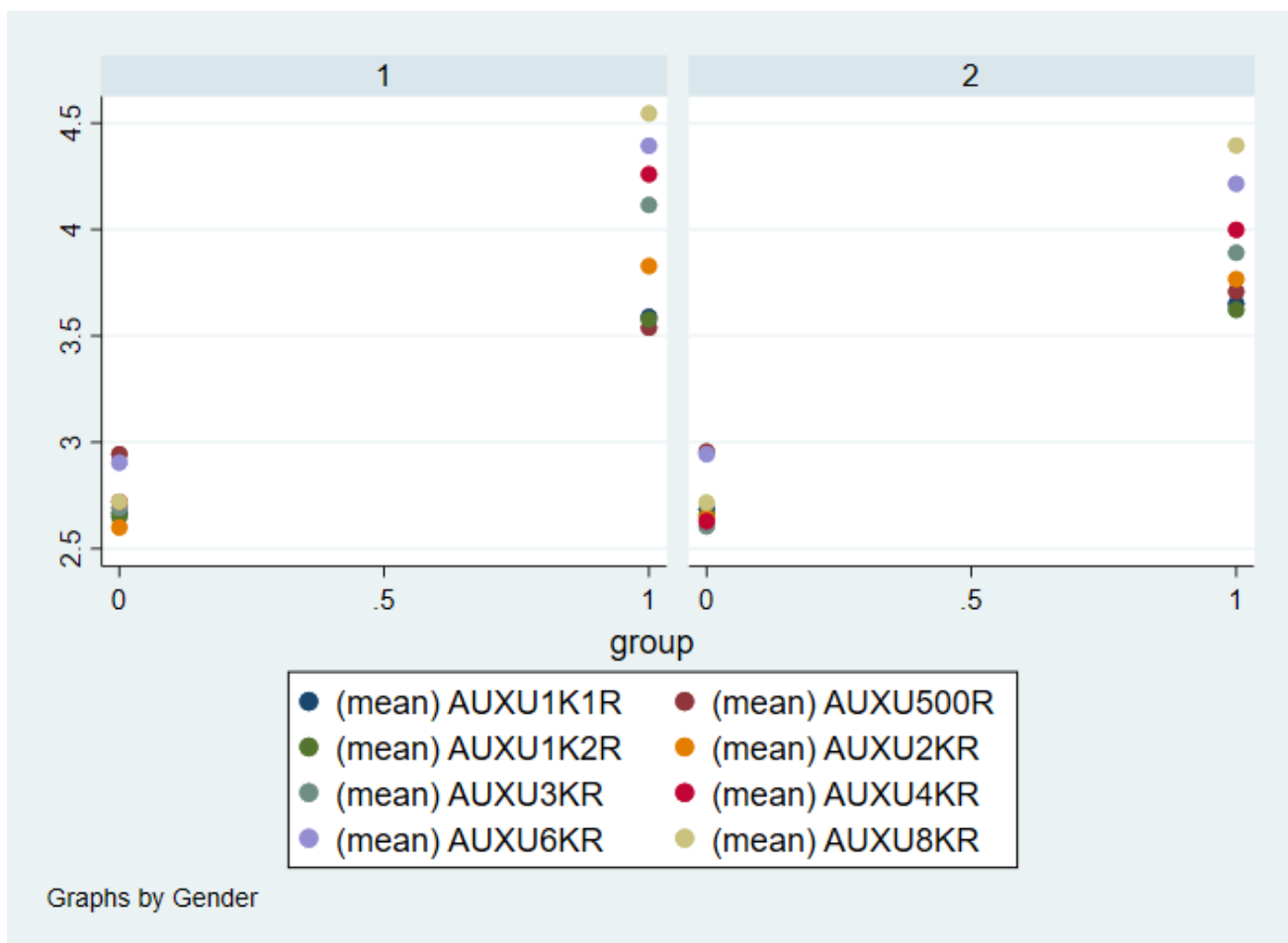


There are no distinct differences between pictures, so I think left ear tests and right ear tests have the same results.

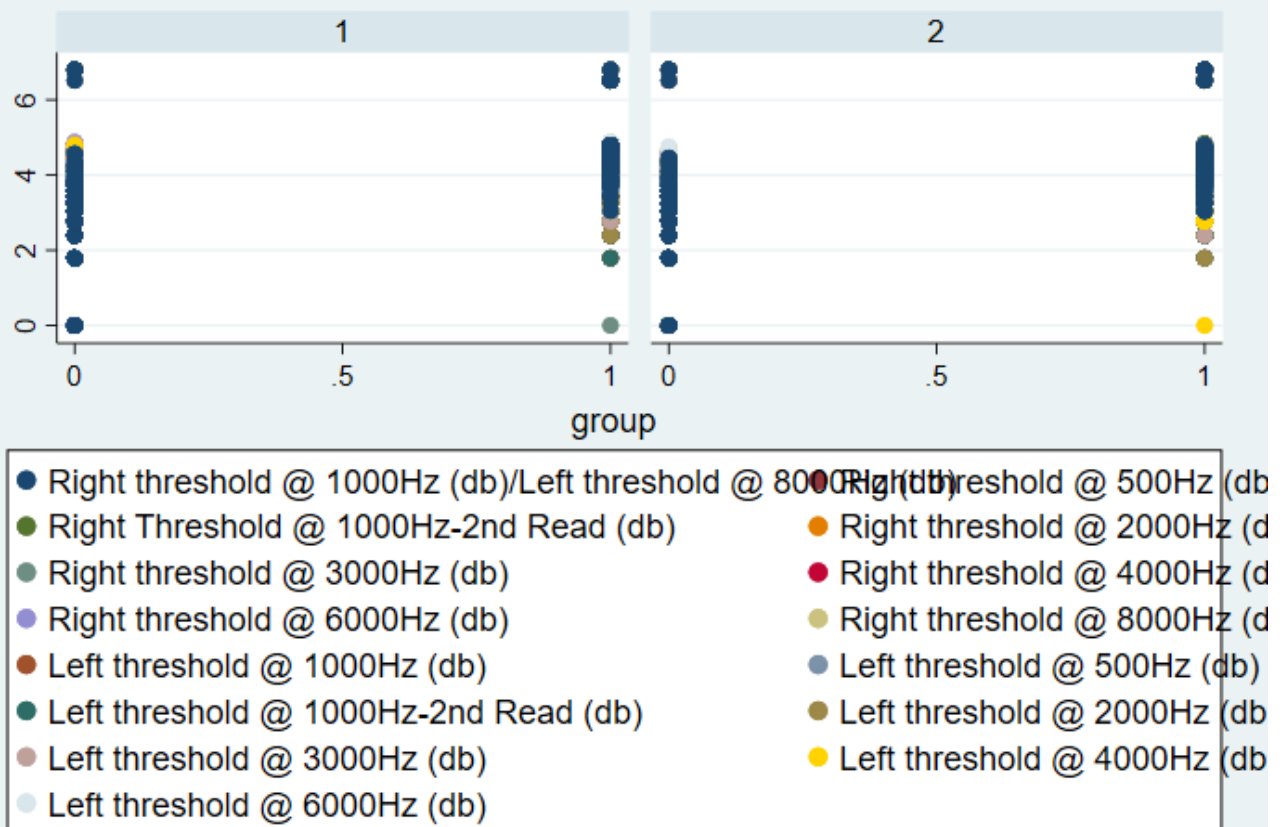
Furthermore, it can be verified by t test. For example, examine differences of left and right ear threshold by `group`, the null hypothesis cannot be rejected.

c.

For right ear, plot a graph about mean:



And plot a graph about all data:



Graphs by Gender

Neither gender is more prone to age-related hearing loss.