STATS 506 PS4 Question 1

Yunbin Peng

December 6, 2017

Author: Yunbin Peng
Data set:
Audiometry & Demographic from NHANES 2005-2006 survey
https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/DEMO_D.XPT
https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/AUX_D.XPT
Description:
This script use proc reg and proc mixed function in SAS to investigate how age, age group, gender affect hearing using NHANES survey.

a

```
libname demog xport 'L:\SAS\Homework4\DEMO_D.XPT';
libname aux xport 'L:\SAS\Homework4\AUX_D.XPT';
libname outlib 'L:\SAS\Homework4';


proc copy in=demog out=outlib;
proc copy in=aux out=outlib;

* Load XPT into dataset;
data demog;
    set outlib.DEMO_D;



* Sort dataset by variable SEQN;
proc sort data=demog;
    by SEQN;

* merge audiometry and demographic data set;
data nhanes;
    merge WORK.demog aux.AUX_D;
        by SEQN;

* drop cases without audiometry data;
data nhanes1;
    set nhanes;
    if (AUXU500R ge 150) or (AUXU500R =.) then delete;
    if (AUXU1K1R ge 150) or (AUXU1K1R =.) then delete;
    if (AUXU2KR ge 150) or (AUXU2KR =.) then delete;
    if (AUXU3KR ge 150) or (AUXU3KR =.) then delete;
    if (AUXU4KR ge 150) or (AUXU4KR =.) then delete;
    if (AUXU6KR ge 150) or (AUXU6KR =.) then delete;
    if (AUXU8KR ge 150) or (AUXU8KR =.) then delete;
    if (AUXU500L ge 150) or (AUXU500L =.) then delete;
```

```
        if (AUXU1K1L ge 150) or (AUXU1K1L =.) then delete;
        if (AUXU2KL ge 150) or (AUXU2KL =.) then delete;
        if (AUXU3KL ge 150) or (AUXU3KL =.) then delete;
        if (AUXU4KL ge 150) or (AUXU4KL =.) then delete;
        if (AUXU6KL ge 150) or (AUXU6KL =.) then delete;
        if (AUXU8KL ge 150) or (AUXU8KL =.) then delete;


* save merged data into local folder;
data outlib.nhanes1;
    set nhanes1;

run;
```

b

```
* keep variable seqn age gender test result, drop cases with missing age gender;
data nhanes2;
    set nhanes1;
    keep SEQN RIDAGEYR RIAGENDR AUXU: ;
    if RIDAGEYR =. then delete;
    if RIAGENDR =. then delete;

* drop the 2nd reading for 1K test, rename 1st reading;
data nhanes3(drop= AUXU1K2L AUXU1K2R);
    set nhanes2 (rename=(
    AUXU1K1R=R1000 AUXU1K1L=L1000
    AUXU500R=R500 AUXU500L=L500
    AUXU2KR=R2000 AUXU2KL=L2000
    AUXU3KR=R3000 AUXU3KL=L3000
    AUXU4KR=R4000 AUXU4KL=L4000
    AUXU6KR=R6000 AUXU6KL=L6000
    AUXU8KR=R8000 AUXU8KL=L8000
    RIDAGEYR=AGE RIAGENDR=GENDER));

* create age group indicator;
data nhanes4;
    set nhanes3;
    AGEGROUP=(AGE GT 25);
    GENDER = GENDER - 1;

* reshape wide to long;
proc transpose data=nhanes4 out=datalong;
    var R500 R1000 R2000 R3000 R4000 R6000 R8000
        L500 L1000 L2000 L3000 L4000 L6000 L8000;
    by SEQN GENDER AGE AGEGROUP;

data datalong;
    set datalong (rename=(COL1=RESULT));
    FREQUENCY = input(substr(_name_,2),5.);
    EAR = substr(_name_,1,1);
```

```sas
        drop _name_ _label_;

data datalong;
    retain SEQN GENDER AGE AGEGROUP EAR FREQUENCY RESULT;
    set datalong;
    RIGHTEAR = (EAR = 'R');
    drop EAR;

* save data to local folder;
data datalong;
    retain SEQN GENDER AGE AGEGROUP RIGHTEAR FREQUENCY RESULT;
    set datalong;

data outlib.datalong;
    set datalong;
```

```
c
* filter data;
data R1K;
    set outlib.datalong;
    if RIGHTEAR = 1;
    if FREQUENCY = 1000;

i
* create 3 dummy variables for interaction between gender and age group;
data model1;
    set R1K;
    male_old = ((GENDER = 0) and (AGEGROUP = 1));
    female_young = ((GENDER = 1) and (AGEGROUP = 0));
    female_old = ((GENDER = 1) and (AGEGROUP = 1));

proc reg data=model1;
    model RESULT = male_old female_young female_old;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.39819	0.33716	16.01	<.0001
male_old	1	20.21765	0.66661	30.33	<.0001
female_young	1	-0.14992	0.47503	-0.32	0.7523
female_old	1	23.35958	0.68108	34.30	<.0001

Since coefficient for interaction term female_old is significant (baseline is male_young), there is a significant interaction between age group and gender in determining how well an individual hears.

```
ii
proc reg data=R1K;
    model RESULT = AGE AGEGROUP GENDER;
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5.94926	1.01388	-5.87	<.0001
AGE	Age at Screening Adjudicated - Recode	1	0.70947	0.06266	11.32	<.0001
AGEGROUP		1	-22.11708	3.90941	-5.66	<.0001
GENDER	Gender	1	0.55491	0.40313	1.38	0.1688

Even after controlling for age group, the coefficient of age is still significant.

```
data model3;
    set R1K;
    AGE_AGEGROUP = AGE*AGEGROUP;

proc reg data=model3;
    model RESULT = AGE AGEGROUP AGE_AGEGROUP;
```

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.12948	1.58210	1.98	0.0480
AGE	Age at Screening Adjudicated - Recode	1	0.14155	0.10102	1.40	0.1613
AGEGROUP		1	-57.86386	6.30661	-9.18	<.0001
AGE_AGEGROUP		1	0.91579	0.12804	7.15	<.0001

Since the coefficient of AGE*AGEGROUP (0.9158) is statistically significant, the effect of age is different for different age groups.

d

i

```
* i;
* generate dummy variables;
data datalong1;
    set datalong;
    male_old = ((GENDER = 0) and (AGEGROUP = 1));
    female_young = ((GENDER = 1) and (AGEGROUP = 0));
    female_old = ((GENDER = 1) and (AGEGROUP = 1));
    AGE_AGEGROUP = AGE*AGEGROUP;

proc mixed data = datalong1;
    class male_old female_young female_old RIGHTEAR FREQUENCY;
    model RESULT = male_old female_young female_old;
    random FREQUENCY RIGHTEAR;
```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
male_old	1	37E3	30599.5	<.0001
female_young	1	37E3	26.88	<.0001
female_old	1	37E3	20746.9	<.0001

All the dummy variables representing interaction between gender and age group are significant.

ii

```
proc mixed data=datalong1;
    class RIGHTEAR FREQUENCY;
    model RESULT = AGE AGEGROUP GENDER;
    random FREQUENCY RIGHTEAR;
```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	37E3	1080.79	<.0001
AGEGROUP	1	37E3	26.26	<.0001
GENDER	1	37E3	277.83	<.0001

After controlling for age group and gender, the coefficient of age is still significant.

iii

```
proc mixed data=datalong1;
    class RIGHTEAR FREQUENCY;
    model RESULT = AGE AGEGROUP AGE_AGEGROUP;
    random FREQUENCY RIGHTEAR;
```

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	37E3	9.30	0.0023
AGEGROUP	1	37E3	393.39	<.0001
AGE_AGEGROUP	1	37E3	461.38	<.0001

The variable AGE_AGEGROUP (interaction between age and age group) is significant, suggesting the effect of age is significantly different for young and old age group.