

---

# Data Management and Visualization

## Week 2 assignment

### Frequency distribution

This project is about the influence of parents on the feelings and health situation of children.

The data I use in this research is from the addhealth dataset. At the beginning of the research, frequency table of each variable is made for the rough understand of the distribution.

#### Code:

```
LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
data new; set mydata.addhealth_pds;
proc sort; by AID;/*sorted by AID*/
label AID="respondent ID"
      BIO_SEX="gender"
      H1GI20="grade"
      H1GH1="general health"
      H1GH2="frequency of headache"
      H1GH6="frequency of feeling weak"
      H1GH8="frequency of feeling tired"
      H1GH17="frequency of poor appetite"
      H1GH18="frequency of poor sleep"
      H1GH19="frequency of trouble relaxing"
      H1GH59A="height in feet"
      H1GH59B="height in inch"
      H1GH60="weight"
      H1FS4="feeling yourself as good as others"
      H1FS8="feeling hopeful"
      H1FS11="feeling happy"
      H1FS12="talking less"
      H1WP9="how close with mother"
      H1WP10="how much does your mother care about you"
      H1WP11="how disappointed your mother would be if you cannot graduate"
      H1WP13="how close with father"
      H1WP14="how much does your father care about you"
      H1WP15="how disappointed your father would be if you cannot graduate"
      H1WP17F="have you had a talk about personal problem with mother"
      H1WP18F="have you had a talk about personal problem with father";
/*type ; at the last*/
/*label changes the name of each variable*/
proc freq; tables BIO_SEX H1GI20 H1GH1 H1GH2 H1GH6 H1GH8 H1GH17 H1GH18 H1GH19
```

H1GH59A H1GH60 H1FS4 H1FS8 H1FS11 H1FS12 H1WP9 H1WP10 H1WP11  
H1WP13 H1WP14 H1WP15 H1WP17F H1WP18F;

/\*proc is short for procedure\*/  
run;

Three frequency tables are shown as following as example.

Table 1 the frequency of gender

gender				
BIO_SEX	frequency	percentage	cumulative frequency	cumulative percentage
1	3147	48.39	3147	48.39
2	3356	51.60	6503	99.98
6	1	0.02	6504	100.00

1 represents male.

2 represents female.

6 represents not provided.

Table 2 general health

general health				
H1GH1	frequency	percentage	cumulative frequency	cumulative percentage
1	1847	28.40	1847	28.40
2	2608	40.10	4455	68.50
3	1605	24.68	6060	93.17
4	408	6.27	6468	99.45
5	28	0.43	6496	99.88
6	3	0.05	6499	99.92
8	5	0.08	6504	100.00

1: excellent

2: very good

3: good

4: fair

5: poor

6: refused

8: do not know

Table 3 frequency of headache

frequency of headache				
H1GH2	frequency	percentage	cumulative frequency	cumulative percentage
0	578	8.89	578	8.89
1	4016	61.75	4594	70.63
2	1441	22.16	6035	92.79
3	390	6.00	6425	98.79
4	72	1.11	6497	99.89
6	3	0.05	6500	99.94
8	4	0.06	6504	100.00

0: never

1: just a few times

2: about once a week

3: almost every day

4: every day

6: refused

8: do not know

## Summary

The frequency distribution shows that approximately, half of the respondents are male, and half of them are female, which should be the result of random sampling. 99.45% of the respondents are healthy (at least “fair” for general health). About 84% of respondents have headache for a few time, and 6% feel headache every. Only 8.89% respondents never felt headache.

Even though most people are in healthy situation, headache may still be a problem affecting most people.