Data analysis tools

Week 2 assignment

Chi square test

I analyzed that if the gender and general health are independent to each other or not.

Null hypothesis is that they are independent.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of H1GH1 by BIO_SEX		
		BIO_SEX(gender)	
H1GH1(general health)	1	2	Total
1	995 15.71 55.09 32.39	811 12.80 44.91 24.86	1806 28.51
2	1250 19.73 49.04 40.69	1299 20.51 50.96 39.82	2549 40.24
3	663 10.47 42.42 21.58	900 14.21 57.58 27.59	1563 24.68
4	153 2.42 39.13 4.98	238 3.76 60.87 7.30	391 6.17
5	11 0.17 44.00 0.36	14 0.22 56.00 0.43	25 0.39
Total	3072 48.50	3262 51.50	6334 100.00
Frequency Missing = 3			

Statistics for Table of H1GH1 by BIO_SEX

Statistic	DF	Value	Prob
Chi-Square	4	68.8258	<.0001
Likelihood Ratio Chi-Square	4	69.0839	<.0001
Mantel-Haenszel Chi-Square	1	66.3901	<.0001
Phi Coefficient		0.1042	
Contingency Coefficient		0.1037	
Cramer's V		0.1042	

Effective Sample Size = 6334
Frequency Missing = 3

The output results show p-value is less than 0.05. At 95% confidence level, null hypothesis can be rejected, which means gender and general health are related to each other.

To find out at which level they are related, post hoc test was conducted.

Adjusted p-value is $0.05 \div 10 = 0.005$.

	1	2	3	4	5
1	-	-	-	-	-
2	<.0001	-	-	-	-
3	<.0001	<.0001	-	-	-
4	<.0001	0.0003	0.2384	-	-
5	0.2682	0.6160	0.8739	0.6291	-

For the following pairs, p-value is less than 0.005, which means null hypothesis can be rejected, they are related to each other.

Pairs with p-value less than 0.005:

1 and 2, 1 and 3, 1 and 4, 2 and 3, 2 and 4.

For other pairs, the evidence is not enough to reject the null hypothesis.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	1	2	Total
1	995 22.85 44.32 55.09	1250 28.70 55.68 49.04	2245 51.55
2	811 18.62 38.44 44.91	1299 29.83 61.56 50.96	2110 48.45
Total	1806 41.47	2549 58.53	4355 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	15.5184	<.0001
Likelihood Ratio Chi-Square	1	15.5354	<.0001
Continuity Adj. Chi-Square	1	15.2769	<.0001
Mantel-Haenszel Chi-Square	1	15.5148	<.0001
Phi Coefficient		0.0597	
Contingency Coefficient		0.0596	
Cramer's V		0.0597	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	1	3	Total
1	995 29.53 60.01 55.09	663 19.68 39.99 42.42	1658 49.21
2	811 24.07 47.40 44.91	900 26.71 52.60 57.58	1711 50.79
Total	1806 53.61	1563 46.39	3369 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	53.8626	<.0001
Likelihood Ratio Chi-Square	1	54.0208	<.0001
Continuity Adj. Chi-Square	1	53.3567	<.0001
Mantel-Haenszel Chi-Square	1	53.8466	<.0001
Phi Coefficient		0.1264	
Contingency Coefficient		0.1254	
Cramer's V		0.1264	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	1	4	Total
1	995 45.29 86.67 55.09	153 6.96 13.33 39.13	1148 52.25
2	811 36.91 77.31 44.91	238 10.83 22.69 60.87	1049 47.75
Total	1806 82.20	391 17.80	2197 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	32.8302	<.0001
Likelihood Ratio Chi-Square	1	32.9430	<.0001
Continuity Adj. Chi-Square	1	32.1935	<.0001
Mantel-Haenszel Chi-Square	1	32.8153	<.0001
Phi Coefficient		0.1222	
Contingency Coefficient		0.1213	
Cramer's V		0.1222	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	1	5	Total
1	995 54.34 98.91 55.09	11 0.60 1.09 44.00	1006 54.94
2	811 44.29 98.30 44.91	14 0.76 1.70 56.00	825 45.06
Total	1806 98.63	25 1.37	1831 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	1.2260	0.2682
Likelihood Ratio Chi-Square	1	1.2182	0.2697
Continuity Adj. Chi-Square	1	0.8188	0.3655
Mantel-Haenszel Chi-Square	1	1.2253	0.2683
Phi Coefficient		0.0259	
Contingency Coefficient		0.0259	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1			
		H1GH1(general health)		
BIO_SEX(gender)		2	3	Total
	1	1250 30.40 65.34 49.04	663 16.12 34.66 42.42	1913 46.52
	2	1299 31.59 59.07 50.96	900 21.89 40.93 57.58	2199 53.48
Total		2549 61.99	1563 38.01	4112 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	17.0691	<.0001
Likelihood Ratio Chi-Square	1	17.1096	<.0001
Continuity Adj. Chi-Square	1	16.8041	<.0001
Mantel-Haenszel Chi-Square	1	17.0650	<.0001
Phi Coefficient		0.0644	
Contingency Coefficient		0.0643	
Cramer's V		0.0644	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1			
		H1GH1(general health)		
BIO_SEX(gender)		2	4	Total
	1	1250 42.52 89.09 49.04	153 5.20 10.91 39.13	1403 47.72
	2	1299 44.18 84.52 50.96	238 8.10 15.48 60.87	1537 52.28
Total		2549 86.70	391 13.30	2940 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	13.3404	0.0003
Likelihood Ratio Chi-Square	1	13.4590	0.0002
Continuity Adj. Chi-Square	1	12.9462	0.0003
Mantel-Haenszel Chi-Square	1	13.3359	0.0003
Phi Coefficient		0.0674	
Contingency Coefficient		0.0672	
Cramer's V		0.0674	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1			
		H1GH1(general health)		
BIO_SEX(gender)		2	5	Total
	1	1250 48.56 99.13 49.04	11 0.43 0.87 44.00	1261 48.99
	2	1299 50.47 98.93 50.96	14 0.54 1.07 56.00	1313 51.01
Total		2549 99.03	25 0.97	2574 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	0.2515	0.6160
Likelihood Ratio Chi-Square	1	0.2523	0.6155
Continuity Adj. Chi-Square	1	0.0903	0.7638
Mantel-Haenszel Chi-Square	1	0.2514	0.6161
Phi Coefficient		0.0099	
Contingency Coefficient		0.0099	
Cramer's V		0.0099	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1			
		H1GH1(general health)		
BIO_SEX(gender)		3	4	Total
	1	663 33.93 81.25 42.42	153 7.83 18.75 39.13	816 41.76
	2	900 46.06 79.09 57.58	238 12.18 20.91 60.87	1138 58.24
Total		1563 79.99	391 20.01	1954 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	1.3902	0.2384
Likelihood Ratio Chi-Square	1	1.3970	0.2372
Continuity Adj. Chi-Square	1	1.2583	0.2620
Mantel-Haenszel Chi-Square	1	1.3895	0.2385
Phi Coefficient		0.0267	
Contingency Coefficient		0.0267	
Cramer's V		0.0267	

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	3	5	Total
1	663 41.75 98.37 42.42	11 0.69 1.63 44.00	674 42.44
2	900 56.68 98.47 57.58	14 0.88 1.53 56.00	914 57.56
Total	1563 98.43	25 1.57	1588 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of BIO_SEX by H1GH1		
	H1GH1(general health)		
BIO_SEX(gender)	4	5	Total
1	153 36.78 93.29 39.13	11 2.64 6.71 44.00	164 39.42
2	238 57.21 94.44 60.87	14 3.37 5.56 56.00	252 60.58
Total	391 93.99	25 6.01	416 100.00

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	0.0252	0.8739
Likelihood Ratio Chi-Square	1	0.0251	0.8741
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0252	0.8739
Phi Coefficient		-0.0040	
Contingency Coefficient		0.0040	
Cramer's V		-0.0040	

Statistics for Table of BIO_SEX by H1GH1

Statistic	DF	Value	Prob
Chi-Square	1	0.2333	0.6291
Likelihood Ratio Chi-Square	1	0.2307	0.6310
Continuity Adj. Chi-Square	1	0.0740	0.7857
Mantel-Haenszel Chi-Square	1	0.2328	0.6295
Phi Coefficient		-0.0237	
Contingency Coefficient		0.0237	
Cramer's V		-0.0237	

My code

```
1  LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
2  data new; set mydata.addhealth_pds;
3
4  if H1GI20=97 then delete; if H1GI20=99 then delete; if H1GI20=96 then delete;
5  if H1GI20=98 then delete;
6
7  if H1GH1=6 then H1GH1=.; if H1GH1=8 then H1GH1=.;
8
9  label AID="respondent ID"
10       BIO_SEX="gender"
11       H1GH1="general health";
12
13 proc sort; by AID;
14 /*chi square test*/
15 proc freq; tables BIO_SEX*H1GH1/chisq;
16 run;
17 /*post hoc test*/
18 data com1; set new;
19 if H1GH1=1 or H1GH1=2;
20 proc freq; tables BIO_SEX*H1GH1/chisq;
21 run;
22
23 data com2; set new;
```

```
24 if H1GH1=1 or H1GH1=3;
25 proc freq; tables BIO_SEX*H1GH1/chisq;
26 run;
27
28 data com3; set new;
29 if H1GH1=1 or H1GH1=4;
30 proc freq; tables BIO_SEX*H1GH1/chisq;
31 run;
32
33 data com4; set new;
34 if H1GH1=1 or H1GH1=5;
35 proc freq; tables BIO_SEX*H1GH1/chisq;
36 run;
37
38 data com5; set new;
39 if H1GH1=2 or H1GH1=3;
40 proc freq; tables BIO_SEX*H1GH1/chisq;
41 run;
42
43 data com6; set new;
44 if H1GH1=2 or H1GH1=4;
45 proc freq; tables BIO_SEX*H1GH1/chisq;
46 run;

47
48 data com7; set new;
49 if H1GH1=2 or H1GH1=5;
50 proc freq; tables BIO_SEX*H1GH1/chisq;
51 run;
52
53 data com8; set new;
54 if H1GH1=3 or H1GH1=4;
55 proc freq; tables BIO_SEX*H1GH1/chisq;
56 run;
57
58 data com9; set new;
59 if H1GH1=3 or H1GH1=5;
60 proc freq; tables BIO_SEX*H1GH1/chisq;
61 run;
62
63 data com10; set new;
64 if H1GH1=4 or H1GH1=5;
65 proc freq; tables BIO_SEX*H1GH1/chisq;
66 run;
```