Inferential Analysis
Testing Moderation
in The Context of
Chi-Square
And
Python

AUTHOR:

LUIS ORELLANA ALTAMIRANO

Testing Moderation

Like the first and second analysis, the "AddHealth" [1] was taken. In this time, the relationship between these two variables were taken into account:

12.	And what was your grade in mathematics?	
1552	1	A
1899	2	B
1521	3	C
950	4	D or lower
353	5	didn't take this subject
69	6	took the subject, but it wasn't graded this way
4	96	refused
128	97	legitimate skip
28	98	don't know

11.	You were happy	
172	0	never or rarely
1230	1	sometimes
2690	2	a lot of the time
2397	3	most of the time or all of the time
6	6	refused
9	8	don't know

The outcome of chi square is:

```
H1ED12    1     2     3     4     5    6    96   97   98
H1FS11
1.0      237   320   301   235    73   22    1   33    8
2.0      605   797   680   384   150   22    0   44    8
3.0      687   735   506   280   121   19    0   41    8
6.0        0     2     0     0     0    0    3    1    0
8.0        0     1     0     0     0    1    0    3    4

chi-square value 2941.429079
p value 0.000000
```

It indicates the relationship between grades in mathematics and how happy the students feel. Moreover, the results got through post hoc analysis are as follow:

```
chi-square value 1150.361891
p value 4.31E-249
COMP1v2  1.0   96.0
H1FS11
1.0         237      1
2.0         605      0
3.0         687      0
6.0           0      3
COMP1v2       1.0   96.0
H1FS11
1.0       0.155003  0.25
2.0       0.395683  0.00
3.0       0.449313  0.00
6.0       0.000000  0.75
```

The most significant interaction between these two variables is within the span 1 to 96 of the variable "grade in mathematics". Therefore, the last two levels should not be considered because they do not have a significant association with the students' happiness.

Finally, so as to know how strong is the relationship between grades in mathematics and how happy the students feel, a third variable was added. This new variable is how close is the student and his/her mother. The variables' leves are:

9. If MOM, ask Q.9-12.
[Point to show card 17.] How close do you feel to your {MOTHER/ADOPTIVE MOTHER/ STEPMOTHER/ FOSTER MOTHER/etc.}?

25	1	not at all
156	2	very little
480	3	somewhat
1229	4	quite a bit
4239	5	very much
2	6	refused
370	7	legitimate skip *[no MOM]*

The results of moderation analysis through chi square are:

```
How close do you feel to your mother? = 1
chi-square value 9.729167
p value 0.464567
How close do you feel to your mother? = 2
chi-square value 16.876652
p value 0.262811
How close do you feel to your mother? = 3
chi-square value 8.092535
p value 0.884450
How close do you feel to your mother? = 4
chi-square value 27.534213
p value 0.016393
How close do you feel to your mother? = 5
chi-square value 159.486696
p value 0.000000
```

It's possible to see the strongest relationship when adding this new variable which is level "5", which means a strong feeling of closeness with his/her mother.

To sum up, there exists a significant statically relationship between grades in mathematics and how happy the students feel. In addition, feeling of closeness that students have with his/her mothers is also important. In particular, the closer the students are to their mothers, better grades in mathematics they have and they tend to be happier.

Appendix

The Python code to reproduce this analysis is as below:

```python
import numpy
import pandas
import statsmodels.formula.api as smf
import statsmodels.stats.multicomp as multi
import scipy.stats
import seaborn
import matplotlib.pyplot as plt

data = pandas.read_csv('addhealth_pds.csv', low_memory=False)

data['H1FS11']=data['H1FS11'].replace(0, numpy.nan)
data['H1ED12']=data['H1ED12'].replace(0, numpy.nan)

data['H1FS11'] = pandas.to_numeric(data['H1FS11'], errors='coerce')
data['H1ED12'] = pandas.to_numeric(data['H1ED12'], errors='coerce')

ct=pandas.crosstab(data['H1FS11'], data['H1ED12'])
print (ct)

# column percentages
colsum=ct.sum(axis=0)
colpct=ct/colsum
print(colpct)

# chi-square
cs= scipy.stats.chi2_contingency(ct)
print ('chi-square value %f'%cs[0])
print ('p value %f'%cs[1])

##################Post Hoc Analysis

sub = data.copy()

for a in [1, 2, 3, 4, 5, 6, 7, 8, 9]:
    for b in range((a + 1),10):

        bb = b
        aa = a
        if b == 7:
            bb = 96
        if b == 8:
            bb = 97
        if b == 9:
            bb = 98
        if a == 7:
            aa = 96
```

```python
        if a == 8:
            aa = 97

        print ('————————')
        recode1 = {aa: aa, bb: bb}
        sub['COMP1v2']= data['H1ED12'].map(recode1)

        # contingency table of observed counts
        ct1=pandas.crosstab(sub['H1FS11'], sub['COMP1v2'])

        cs1= scipy.stats.chi2_contingency(ct1)
        if (cs1[0] > 3.84) & (cs1[1] < 0.003) :
            print ('chi-square value %f'%cs1[0])
            print ('p value %.2E'%cs1[1])

            print (ct1)

            # column percentages
            colsum=ct1.sum(axis=0)
            colpct=ct1/colsum
            print(colpct)
```

"' The Best one
————————
chi-square value 1150.361891
p value 4.31E-249
```
COMP1v2  1.0    96.0
H1FS11
1.0       237     1
2.0       605     0
3.0       687     0
6.0         0     3
COMP1v2      1.0    96.0
H1FS11
1.0       0.155003  0.25
2.0       0.395683  0.00
3.0       0.449313  0.00
6.0       0.000000  0.75
```
————————
"'

```python
################## Moderation Analysis
for a in range(1, 6):

    print ('How close do you feel to your mother? = %d'%a)
    sub2=data[(data['H1WP9']==a)]

    ct=pandas.crosstab(sub2['H1FS11'], sub2['H1ED12'])

    colsum=ct.sum(axis=0)
    colpct=ct/colsum
```

```python
# chi-square
cs= scipy.stats.chi2_contingency(ct)
print ('chi-square value %f'%cs[0])
print ('p value %f'%cs[1])
```