# Data pre-processing

0.1 Add a binary variable Y to the data set: Y = 1 for individual who thinks he/she has psychological problems, and Y = 0 otherwise. In another word, every observation from group A, B, and C has Y = 1; Y = 0 for those from group AH.

0.2 Divided the total score into three levels (denoted by t). t=0 for total score equal to or lower than the health level, 208. t = 1 for total score higher than 208 but lower than 208 plus on standard deviation (52). t = 2 otherwise.

# Sample Characteristics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Group A, N=3151 | Group B, N=1233 | Group C, N=2246 | Group AH (stratified), N=3029 |
| Demographics | Mean ± SD or N (%) | Mean ± SD or N (%) | Mean ± SD or N (%) | Mean ± SD or N (%) |
| Age(years) | 31.49 ± 10.23 | 33.00 ± 11.45 | 35.49 ± 14.91 | 33.01 ± 7.85 |
| Sex |  |  |  |  |
| 1 | 1670 (53.0) | 649 (52.6) | 1487 (66.2) | 1384 (45.7) |
| 2 | 1481 (47.0) | 584 (47.4) | 759 (33.8) | 1645 (54.3) |
| Education |  |  |  |  |
| 0 |  | 40 (3.2) | 92 (4.1) |  |
| 1 |  | 91 (7.4) | 202 (9.0) | 40 (1.3) |
| 2 | 1220 (38.7) | 529 (42.9) | 882 (39.3) | 1599 (52.8) |
| 3 | 643 (20.4) | 505 (41.0) | 917 (40.8) | 1182 (39.0) |
| 4 | 1288 (40.9) | 68 (5.5) | 153 (6.8) | 208 (6.9) |
| Marriage |  |  |  |  |
| 1 | 1828 (58.0) | 730 (59.2) | 1347 (60.0) | 2029 (67.0) |
| 2 | 1290 (40.9) | 469 (38.0) | 872 (38.8) | 945 (31.2) |
| 3 |  |  | 10 (0.4) |  |
| 4 | 33 (1.1) | 34 (2.8) | 17 (0.8) | 55 (1.8) |
| SCL90 Result |  |  |  |  |
| Total Score | 205.38 ± 63.33 | 198.21 ± 56.71 | 177.78 ± 64.92 | 145.72 ± 52.10 |
| Somatization | 1.99 ± 0.76 | 1.98 ± 0.69 | 1.8 ± 0.74 | 1.69 ± 0.67 |
| Obsessive.compulsive | 2.52 ± 0.85 | 2.51 ± 0.79 | 2.17 ± 0.81 | 1.85 ± 0.67 |
| Interpersonal.sensitivity | 2.29 ± 0.87 | 2.24 ± 0.86 | 1.99 ± 0.86 | 1.6 ± 0.63 |
| Depression | 2.58 ± 0.91 | 2.45 ± 0.85 | 2.17 ± 0.94 | 1.68 ± 0.68 |
| Anxiety | 2.52 ± 0.9 | 2.34 ± 0.8 | 2.13 ± 0.91 | 1.65 ± 0.65 |
| Hostility | 2.21 ± 0.92 | 2.14 ± 0.8 | 1.91 ± 0.88 | 1.56 ± 0.68 |
| Phobic.anxiety | 1.89 ± 0.83 | 1.75 ± 0.72 | 1.74 ± 0.78 | 1.36 ± 0.53 |
| Paranoid.ideation | 2.07 ± 0.85 | 2.04 ± 0.8 | 1.79 ± 0.81 | 1.51 ± 0.62 |
| Psychoticism | 2.1 ± 0.77 | 1.98 ± 0.71 | 1.79 ± 0.73 | 1.48 ± 0.56 |
| Additional.items | 2.42 ± 0.82 | 2.39 ± 0.75 | 2.1 ± 0.83 | 1.63 ± 0.64 |

# 2. A， B， C三地人群心理因素的共同特征

## 2.1 Analysis the locations separately

Use the chi square statistics to test whether the total score is independent with the basic identity factors (Age, Sex, Education, and Marriage):

H0: the total score level and factor X are independent.

H1: the total score level and factor X are not independent. (X = Age, Sex, Education, or Marriage)

P value table for Group A: