**Chapter 13: PCA and Factor Analysis**

**PCA**

Open the Girls.csv data file in Jamovi. This is an imaginary study about teenage girls' reluctance to participate in school sports, based on interview ratings and surveys. At this stage, assume that we have no underlying theory.

The variable headers are:

interest, agility, contact, strength, clothes, looks, internal, speed, teacher, masculine, harassment

Conduct a Principal Components Analysis (PCA) to reduce the number of variables that describe the characteristics.

Describe the assumption test results and the PCA results.





There are two components as seen on the component loadings table. Since both have more than 3 variables, it is valid to keep both components since they are more likely to be stable.

| Summary |
| --- |
|  |  |  |  |  |  |  |  |
| **Component** | **SS Loadings** | **% of Variance** | **Cumulative %** |
| 1 |  | 4.67 |  | 42.5 |  | 42.5 |  |
| 2 |  | 4.59 |  | 41.7 |  | 84.2 |  |
|  |

| Bartlett's Test of Sphericity |
| --- |
|  |  |  |  |  |  |
| **χ²** | **df** | **p** |
| 408 |  | 55 |  | < .001 |  |
|  |

 The Bartlett’s test result is significant since the *p* value is less than 0.05. This means that we reject the null hypothesis that all the correlations tested simultaneously are ’not statistically different from zero’. This result is necessary to carry on with the interpretation of the analysis’ results.



The measures of sampling adequacy for each variable should be at least .5, and the nearer to 1 the better. Here, we see that the overall statistic is over the minimum for all variables. Thus, we don’t need to remove any of them to continue.

**Scree Plot**



The scree plot supports the result of having two components. There are only two points above the scree (the rocks at the bottom of the hill). Also, the eigenvalues from the third component onwards are quite low.

**EFA**

Use the Job Burnout.csv file.

This is a survey of 99 teachers using the Maslach-Burnout Inventory General Survey (MBI-GS). The survey questions are as follows:

Question 1 I feel emotionally drained from my work.

Question 2 I feel used up at the end of the workday.

Question 3 I feel tired when I get up in the morning and have to face another day on the job.

Question 4 Working all day is really a strain for me.

Question 5 I can effectively solve the problems that arise in my work.

Question 6 I feel burned out from my work.

Question 7 I feel I am making an effective contribution to what this organization does.

Question 8 I have become less interested in my work since I started this job.

Question 9 I have become less enthusiastic about my work.

Question 10 In my opinion, I am good at my job.

Question 11 I feel exhilarated when I accomplish something at work.

Question 12 I have accomplished many worthwhile things in this job.

Question 13 I just want to do my job and not be bothered.

Question 14 I have become more cynical about whether my work contributes anything.

Question 15 I doubt the significance of my work.

Question 16 At my work, I feel confident that I am effective at getting things done.

The MBI-GS suggests that job burnout comprises of three major subscales namely exhaustion, cynicism, and professional efficacy. Based on this scale, questions 1,2,3,4 and 6 correspond to exhaustion, questions 8,9,13, 14, and 15 correspond to cynicism, and questions 5,7,10,11, and 12 correspond to professional efficacy

Conduct an exploratory factor analysis to see if the data will show the same three categories of underlying factors.



| Bartlett's Test of Sphericity |
| --- |
|  |  |  |  |  |  |
| **χ²** | **df** | **p** |
| 809 |  | 120 |  | < .001 |  |
|  |

  The Bartlett’s test result is significant since the p-value is less than 0.05. This means that we reject the null hypothesis that all the correlations tested simultaneously are ’not statistically different from zero’. This result is necessary to carry on with the interpretation of the analysis’ results.

| KMO Measure of Sampling Adequacy |
| --- |
|  |  |  |  |
|  | **MSA** |
| Overall |  | 0.810 |  |
| Question1 |  | 0.830 |  |
| Question 2 |  | 0.878 |  |
| Question 3 |  | 0.885 |  |
| Question 4 |  | 0.922 |  |
| Question 5 |  | 0.783 |  |
| Question 6 |  | 0.874 |  |
| Question 7 |  | 0.737 |  |
| Question 8 |  | 0.804 |  |
| Question 9 |  | 0.792 |  |
| Question 10 |  | 0.806 |  |
| Question 11 |  | 0.731 |  |
| Question 12 |  | 0.805 |  |
| Question 13 |  | 0.561 |  |
| Question 14 |  | 0.653 |  |
| Question 15 |  | 0.725 |  |
| Question 16 |  | 0.816 |  |
|  |

The measures of sampling adequacy for each variable should be at least .5, and the nearer to 1 the better. Here, we see that the overall statistic is over the minimum for all variables. So we don’t need to remove any of them to continue.

**Scree Plot**



The scree plot suggests three factors. The factor loadings indicate similarly, using the default identification method (parallel analysis).

| Factor Loadings |
| --- |
|  | **Factor** |  |
|  | **1** | **2** | **3** | **Uniqueness** |
| Question1 |  | 0.769 |  |   |  |   |  | 0.433 |  |
| Question 2 |  | 0.829 |  |   |  |   |  | 0.396 |  |
| Question 3 |  | 0.811 |  |   |  |   |  | 0.313 |  |
| Question 4 |  | 0.786 |  |   |  |   |  | 0.338 |  |
| Question 5 |  |   |  | 0.624 |  |   |  | 0.620 |  |
| Question 6 |  | 0.902 |  |   |  |   |  | 0.175 |  |
| Question 7 |  |   |  | 0.732 |  |   |  | 0.466 |  |
| Question 8 |  | 0.632 |  |   |  |   |  | 0.446 |  |
| Question 9 |  | 0.625 |  |   |  |   |  | 0.472 |  |
| Question 10 |  |   |  | 0.682 |  |   |  | 0.497 |  |
| Question 11 |  |   |  | 0.431 |  |   |  | 0.774 |  |
| Question 12 |  |   |  | 0.728 |  |   |  | 0.404 |  |
| Question 13 |  |   |  |   |  | 0.497 |  | 0.664 |  |
| Question 14 |  |   |  |   |  | 0.877 |  | 0.191 |  |
| Question 15 |  |   |  |   |  | 0.307 |  | 0.771 |  |
| Question 16 |  |   |  | 0.814 |  |   |  | 0.330 |  |
| Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation |
|  |

Although there is some discrepancy from the original grouping of questions, there is still a resemblance to the subscales as suggested by the MBI-GS.

| Model Fit Measures |
| --- |
|  | **RMSEA 90% CI** |  | **Model Test** |
| **RMSEA** | **Lower** | **Upper** | **TLI** | **BIC** | **χ²** | **df** | **p** |
| 0.0889 |  | 0.0530 |  | 0.105 |  | 0.888 |  | -223 |  | 122 |  | 75 |  | < .001 |  |
|  |

 To indicate a good fit, the RMSEA should be as small as possible and the TLI should be high. These were satisfied since RMSEA is only 0.089 and TLI is 0.88.