Vivian Ellis

Women’s Track Records – DSCI 449 Final

6/11/2018

I will be examining the women’s track dataset, this includes the 100, 200, 400, 800, 1500, 3000-meter dash and the marathon time completion as well as the associated country. The women’s track dataset has 55 observations. Factor analysis will help to understand if these attributes could be grouped and possibly reduced. In particular, I will use exploratory factor analysis thus there will be no hypothesis since I do not know how the factors behave. First, I will be examining the simple statistics and correlation matrix for the 7 metric variables: 100, 200, 400, 800, 1500, 3000-meter dash and the marathon time completion.

| **Table 1 – Simple Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **\_100m** | 55 | 11.61855 | 0.45221 | 639.02000 | 10.79000 | 12.90000 |
| **\_200m** | 55 | 23.64164 | 1.11106 | 1300 | 21.71000 | 27.10000 |
| **\_400m** | 55 | 53.40582 | 2.67834 | 2937 | 47.99000 | 60.40000 |
| **\_800m** | 55 | 2.07636 | 0.10822 | 114.20000 | 1.89000 | 2.33000 |
| **\_1500m** | 55 | 4.32545 | 0.33243 | 237.90000 | 3.87000 | 5.81000 |
| **\_3000m** | 55 | 9.44764 | 0.82434 | 519.62000 | 8.45000 | 13.04000 |
| **marathon** | 55 | 173.25327 | 30.42954 | 9529 | 142.72000 | 306.00000 |

Table 1 shows that the 100-meter dash has an average completion time of 11.61 seconds and the fastest time of 10.79 seconds with the longest time of 12.90 seconds. The 200-meter dash and the 400-meter dash have an average completion time of 23.64 seconds and 53.40 seconds respectively. The middle-distant runs; 800-meter, 1500-meter and 3000-meter have an average completion time of 2.07 seconds, 4.32 seconds, and 9.44 seconds with a standard deviation of .108 seconds, .332 seconds and .824 seconds respectively.

The Pearson’s Corelation matrix as seen in table 2 has 21 correlations that are significant at the p-value < .0001 level.

| **Table 2 – Pearson Correlation Coefficients, N = 55  Prob > |r| under H0: Rho=0** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **\_100m** | **\_200m** | **\_400m** | **\_800m** | **\_1500m** | **\_3000m** | **marathon** |
| **\_100m** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.95279 | | <.0001 | | |  | | --- | | 0.83469 | | <.0001 | | |  | | --- | | 0.72769 | | <.0001 | | |  | | --- | | 0.72837 | | <.0001 | | |  | | --- | | 0.74170 | | <.0001 | | |  | | --- | | 0.68634 | | <.0001 | |
| **\_200m** | |  | | --- | | 0.95279 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.85696 | | <.0001 | | |  | | --- | | 0.72406 | | <.0001 | | |  | | --- | | 0.69836 | | <.0001 | | |  | | --- | | 0.70987 | | <.0001 | | |  | | --- | | 0.68557 | | <.0001 | |
| **\_400m** | |  | | --- | | 0.83469 | | <.0001 | | |  | | --- | | 0.85696 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.89841 | | <.0001 | | |  | | --- | | 0.78784 | | <.0001 | | |  | | --- | | 0.77764 | | <.0001 | | |  | | --- | | 0.70542 | | <.0001 | |
| **\_800m** | |  | | --- | | 0.72769 | | <.0001 | | |  | | --- | | 0.72406 | | <.0001 | | |  | | --- | | 0.89841 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.90161 | | <.0001 | | |  | | --- | | 0.86357 | | <.0001 | | |  | | --- | | 0.77929 | | <.0001 | |
| **\_1500m** | |  | | --- | | 0.72837 | | <.0001 | | |  | | --- | | 0.69836 | | <.0001 | | |  | | --- | | 0.78784 | | <.0001 | | |  | | --- | | 0.90161 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.96917 | | <.0001 | | |  | | --- | | 0.87793 | | <.0001 | |
| **\_3000m** | |  | | --- | | 0.74170 | | <.0001 | | |  | | --- | | 0.70987 | | <.0001 | | |  | | --- | | 0.77764 | | <.0001 | | |  | | --- | | 0.86357 | | <.0001 | | |  | | --- | | 0.96917 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.89984 | | <.0001 | |
| **marathon** | |  | | --- | | 0.68634 | | <.0001 | | |  | | --- | | 0.68557 | | <.0001 | | |  | | --- | | 0.70542 | | <.0001 | | |  | | --- | | 0.77929 | | <.0001 | | |  | | --- | | 0.87793 | | <.0001 | | |  | | --- | | 0.89984 | | <.0001 | | |  | | --- | | 1.00000 | |  | |

Next, I will find the Measure of Sampling Adequacy (MSA) which is a summary of how small the partial correlations are relative to the ordinary correlations and the partial correlations which are the correlations between variables controlling for all other variables. There is 3 high partial correlations, see table 3 below, when we define a partial correlation above .7 as high.

| **Table 3 – Partial Correlations Controlling all other Variables** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **\_100m** | **\_200m** | **\_400m** | **\_800m** | **\_1500m** | **\_3000m** | **marathon** |
| **\_100m** | 1.00000 | 0.82249 | 0.00934 | -0.06913 | 0.06281 | 0.14757 | -0.17603 |
| **\_200m** | 0.82249 | 1.00000 | 0.39107 | -0.13619 | -0.01585 | -0.11515 | 0.25530 |
| **\_400m** | 0.00934 | 0.39107 | 1.00000 | 0.72138 | -0.22668 | 0.12859 | -0.11041 |
| **\_800m** | -0.06913 | -0.13619 | 0.72138 | 1.00000 | 0.52406 | -0.16644 | 0.01749 |
| **\_1500m** | 0.06281 | -0.01585 | -0.22668 | 0.52406 | 1.00000 | 0.75945 | 0.06083 |
| **\_3000m** | 0.14757 | -0.11515 | 0.12859 | -0.16644 | 0.75945 | 1.00000 | 0.40276 |
| **marathon** | -0.17603 | 0.25530 | -0.11041 | 0.01749 | 0.06083 | 0.40276 | 1.00000 |

The overall MSA = 0.83796428 is good since it is >.80. All the variable have a MSA >.50, refer to table 4, all of the variables are suitable for factor analysis.

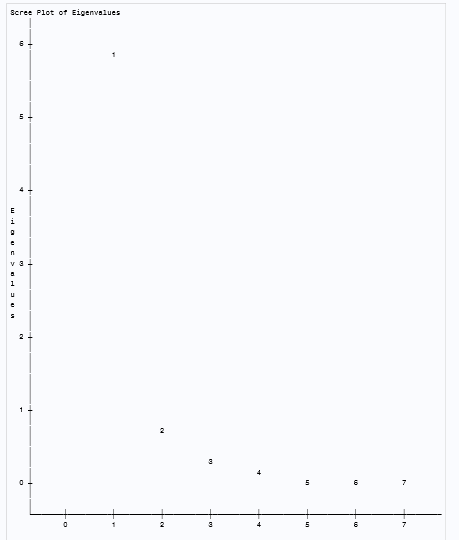
| **Table 4 – Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.83796428** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **\_100m** | **\_200m** | **\_400m** | **\_800m** | **\_1500m** | **\_3000m** | **marathon** |
| 0.83316295 | 0.79655219 | 0.84017928 | 0.82632533 | 0.82046989 | 0.83541950 | 0.92961468 |

To determine the number of factors to extract, we will use table 5 and the eigenvalue criterion to extract the number of factors in which the eigenvalues are > 1. One factors will be retained by the NFACTOR criterion, this means that all 7 variables could be grouped together such that there is only one factor.

| **Table 5 – Eigenvalues of the Correlation Matrix: Total = 7 Average = 1** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 5.80568576 | 5.15204024 | 0.8294 | 0.8294 |
| **2** | 0.65364552 | 0.35376309 | 0.0934 | 0.9228 |
| **3** | 0.29988243 | 0.17440494 | 0.0428 | 0.9656 |
| **4** | 0.12547749 | 0.07166058 | 0.0179 | 0.9835 |
| **5** | 0.05381692 | 0.01476763 | 0.0077 | 0.9912 |
| **6** | 0.03904928 | 0.01660668 | 0.0056 | 0.9968 |
| **7** | 0.02244260 |  | 0.0032 | 1.0000 |

To verify that the number of factors to use, the scree test will be implemented. Notice the ‘elbow’ in figure 1 occurs at 1.

Figure 1 : Scree Test



The unrotated factor solution each variable to loads high on factor1, it is not possible to perform a rotation with 1 factor.

| **Table 6 – Factor Pattern** | |
| --- | --- |
|  | **Factor1** |
| **\_100m** | 0.88755 |
| **\_200m** | 0.88034 |
| **\_400m** | 0.91949 |
| **\_800m** | 0.92660 |
| **\_1500m** | 0.93755 |
| **\_3000m** | 0.93697 |
| **marathon** | 0.88430 |

Communality is the total amount of variance an original variable shares with all other variables included in the analysis and all communalities are greater than 0.5.

| **Table 7 – Final Communality Estimates: Total = 5.805686** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **\_100m** | **\_200m** | **\_400m** | **\_800m** | **\_1500m** | **\_3000m** | **marathon** |
| 0.78775152 | 0.77500668 | 0.84546143 | 0.85857844 | 0.87899191 | 0.87791744 | 0.78197833 |

The reliability of scale1 using the standardized Cronbach Coefficient Alpha is .96, recall that the closer to 1 the more reliable the scale. The reliability measures for all 7 variables is at least .95.

Further work with this dataset would include exploring the multicollinearity possibly present in this dataset.

**Appendix – Factor Analysis on Women’s Track Data**

| **Cronbach Coefficient Alpha** | |
| --- | --- |
| **Variables** | **Alpha** |
| Raw | 0.257778 |
| Standardized | 0.965588 |

| **Cronbach Coefficient Alpha with Deleted Variable** | | | | |
| --- | --- | --- | --- | --- |
| **Deleted Variable** | **Raw Variables** | | **Standardized Variables** | |
| **Correlation with Total** | **Alpha** | **Correlation with Total** | **Alpha** |
| **\_100m** | 0.731720 | 0.247162 | 0.849082 | 0.962149 |
| **\_200m** | 0.723547 | 0.221010 | 0.839874 | 0.962838 |
| **\_400m** | 0.728805 | 0.155146 | 0.889083 | 0.959136 |
| **\_800m** | 0.817802 | 0.260358 | 0.896250 | 0.958592 |
| **\_1500m** | 0.897388 | 0.248931 | 0.910923 | 0.957475 |
| **\_3000m** | 0.914714 | 0.223706 | 0.910599 | 0.957500 |
| **marathon** | 0.793427 | 0.772649 | 0.841291 | 0.962732 |

|  |  |
| --- | --- |
| **1 Variables:** | scale1 |

| **Simple Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **scale1** | 55 | 26.41026 | 4.39167 | 1453 | 21.93000 | 45.53429 |

| **Pearson Correlation Coefficients, N = 55  Prob > |r| under H0: Rho=0** | |
| --- | --- |
|  | **scale1** |
| **scale1** | |  | | --- | | 1.00000 | |  | |