**Andrew Hillard**

**Data 2**

**Homework 6**

**Problem 2**

1. The sphericity test is testing whether the variances of the differences between all possible pairs of groups are equal. In this example, this is testing whether the variances of the differences between all possible pairs of times (i.e. 8, 10, 12, and 14) are equal. The test is important because sphericity is an assumption of repeated-measures ANOVA and determines whether we can use univariate testing or not.

The sphericity test for the Pothoff and Roy data gives a p-value of 0.1997. We therefore fail to reject the null hypothesis and the univariate approach is appropriate, which means we can treat the independent variable as “time” with four levels. Alternatively, we could use the multivariate approach as it does not require the sphericity assumption.

1. The parallel profile test produces a Wilks’ Lambda value of 0.73988, a F Value of 2.70, numerator and denominator degrees of freedom of 3 and 23, and a p-value of .0696. The conclusion is that the profiles are parallel at the 5% significance level.

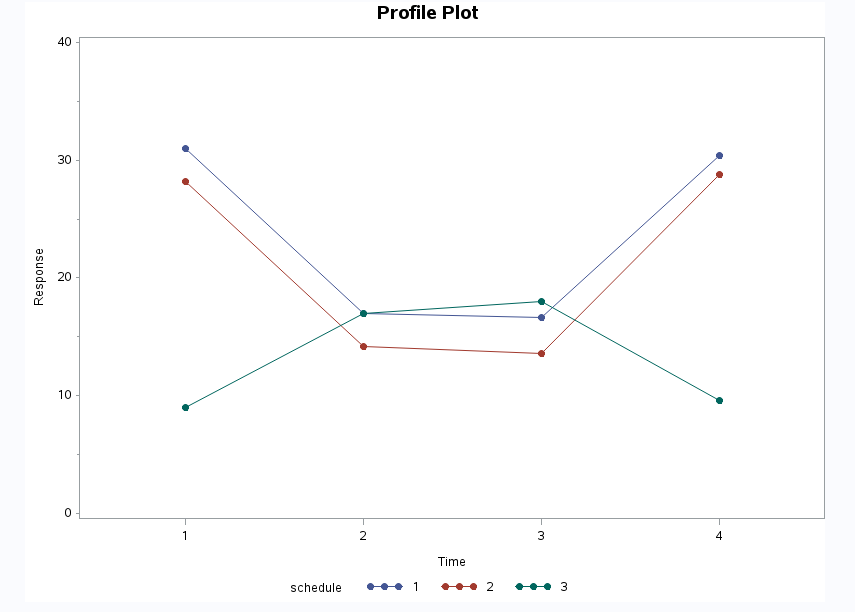
The coincidental profiles test produces a Wilks’ Lambda value of 0.72903, a F Value of 9.29, numerator and denominator degrees of freedom of 1 and 25, and a p-value of .0054. The conclusion is that the profiles are not coincidental.

The horizontal profiles test produces a Wilks’ Lambda value of 0.19479, a F Value of 31.69, numerator and denominator degrees of freedom of 3 and 23, and a p-value of <.0001. The conclusion is that the profiles are not horizontal.

1. The age\*gender interaction has a p-value of 0.0781 and is therefore not significant at the 5% significance level. The age variable has a p-value of <.0001 and is therefore significant at the 5% level. The conclusion is that while there is no age gender interaction, age does have a significant effect on the response measurement.
2. The Greenhouse-Geisser and Huynh-Feldt adjustments are used to adjust the degrees of freedom in the F distribution when the assumption of sphericity is violated. The adjustments correct the F distribution in order to decrease Type 1 Error which is inflated when sphericity is violated. In this problem, the Greenhouse-Geisser adjustment is 0.8672 and the Huynh-Feldt adjustment is 0.9769.
3. The gender\*age interaction is not significant with a p-value of 0.0781. The gender main effect is significant with a p-value of 0.0054. The age main effect is also significant with a p-value of <.0001. The gender\*age interaction and age conclusions are the same as part c.

**Problem 3**

1. Please see below for the profile plot. The profile plot suggests that reinforcement schedule 1 and 2 have similar response times for all conditions. Reinforcement schedule 3 has similar response times to reinforcement schedule 1 and 2 for conditions 2 and 3, but different response times for conditions 1 and 4. The result of the plot therefore suggests that there is a time and reinforcement schedule interaction.



1. Yes, there is an overall reinforcement schedule effect. The overall reinforcement schedule effect was tested using MANOVA and the Wilks’ Lambda value was 0.1384, with an F Value of 3.80, numerator and denominator degrees of freedom equal to 8 and 18, and a p-value of 0.0089.
2. I tested the hypothesis that schedule 1 and 2 behave the same by first running a parallel profile test. The test produced a p-value of 0.9976; therefore, schedule 1 and 2 are parallel profiles at a 5% significance level. I then tested the hypothesis that schedule 1 and 2 are coincidental profiles and found a p-value of 0.4595; therefore, schedule 1 and 2 are coincidental profiles at a 5% significance level, aka schedule 1 and 2 behave the same.

I then tested whether schedule 3 behaved differently than schedule 1 and 2. Because schedule 1 and 2 were found to behave similarly, I only need to show that schedule 3 behaves differently than either schedule 1 or schedule 2. I ran a parallel profile test comparing schedule 1 and schedule 3 and found a p-value of 0.0069. I also ran a parallel profile test comparing schedule 2 and schedule 3 and found a p-value of 0.0064. Therefore, schedule 3 behaves differently than schedule 1 and schedule 2.

**Problem 4**

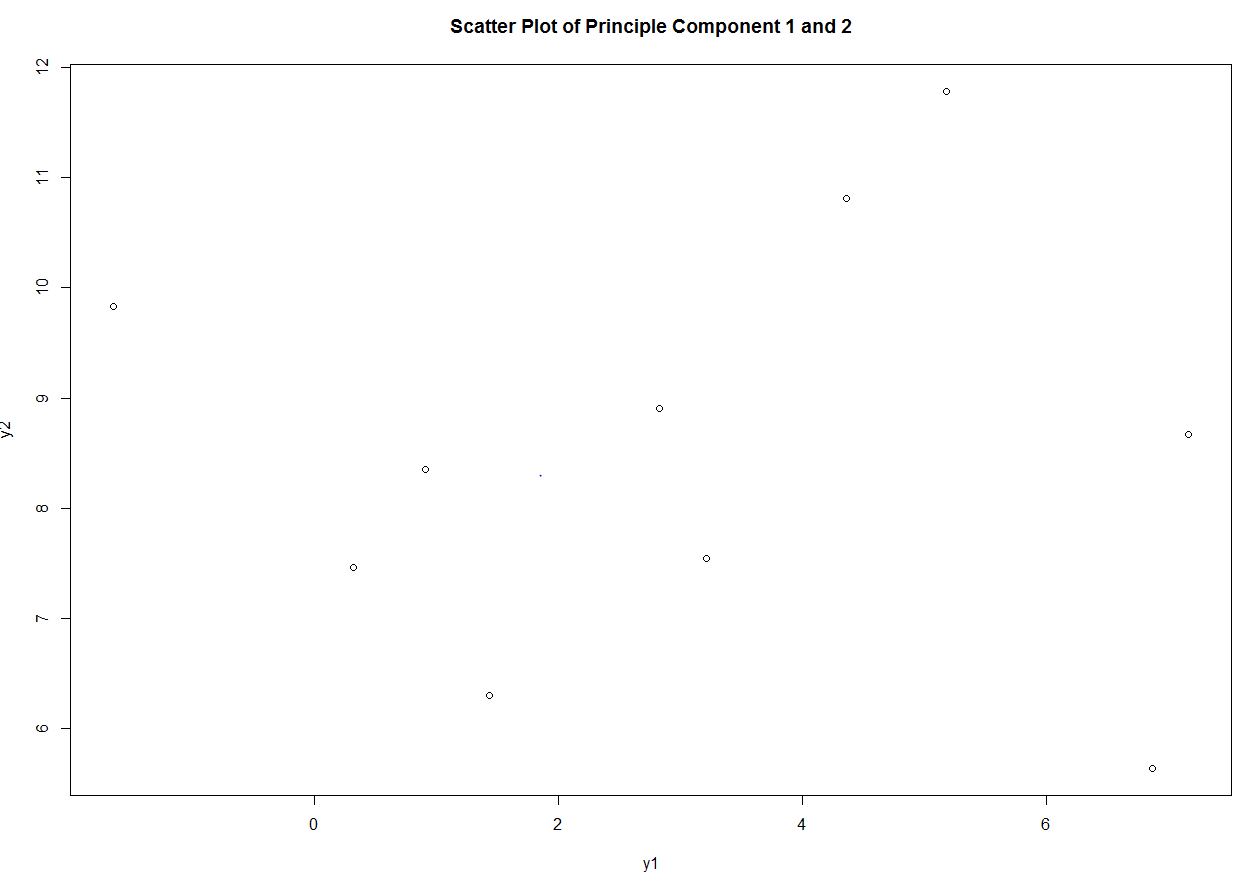
1. The R output for the estimated covariance matrix for the three traits is given below.



1. 12.7
2. and
3. 0.6515
4. 0.9410
5. Used R with x matrix and loading coefficients to find y1 and y2 given below.
6. The loading coefficients for the first two principal components are given below.

The first principal component has a slightly negative correlation with trait 1 and trait 2 and an extremely strong positive correlation with trait 3. The second principal component has strong positive correlations with trait 1 and trait 2, which is the reverse of principal component 1, and a slight positive correlation for trait 3.Principal component 1 and 2 complement each other well in terms of the direction of correlation for traits 1 through 3.

1. The scatterplot seems to show a slight positive correlation between y1 and y2 but it’s hard to tell. The plot could also be interpreted as random.



1. The estimated correlation matrix is given as follows.
2. and

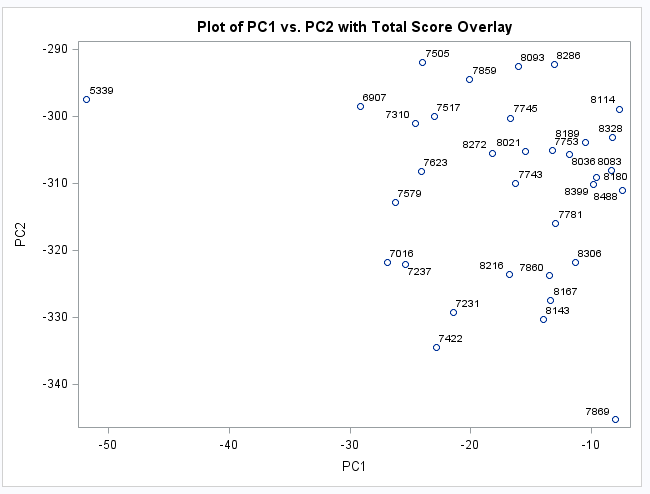
The first principal component accounts for 0.5896 of the total variation. The second principal component accounts for 0.3090 of the total variation. The first and second principal component combined account for 0.8986 of the total variation.

**Problem 5**

1. It makes sense to consider the correlation matrix in this example since we are comparing performance of different athletic events and being good at a certain type of athletic event (e.g. one hundred meter dash) would intuitively make you better at a similar event (e.g. four hundred meter dash). Or, alternatively, being good at one event (e.g. the discus) would intuitively make you worse at another event (e.g. four hundred meter dash) because they require a different body type.
2. The amount of variation accounted in the first 2 PCs is 0.7103.
3. It is reasonable to only consider the first 2 PCs since they account for over 70% of the variability and the scree plot shows a significant drop between the 2nd and 3rd PC and levels off after the 3rd PC.
4. An interpretation of Factor 1 is that it describes the latent variable of general athletic ability. Factor 1 is positively correlated to all events. You should do well in all events the more athletic you are.

An interpretation of Factor 2 is that it describes the latent variable of upper body strength. Factor 2 goes from being positively correlated to negatively correlated as you move from events that don’t require upper body strength (e.g. one hundred, four hundred, and fifteen hundred meters) to events that require upper body strength (e.g. shotput, discus, and javelin). This makes sense with what is typically observed at track and field competitions. The lighter you are, the better you will do in pure running events. The heavier or more muscular you are, the better you will do in events that require more muscle mass like shotput, javelin, and discus.

1. It appears that the top scores appear to be clustered on the right side of the plot. The further you go left, the lower the total score. In other words, the total score appears to be positively correlated with PC1. The spread of scores across the range of PC2 doesn’t show as clear a trend. There are high scores at the top and bottom of the plot.



1. Principal component 1 and total score have a correlation of 0.92059. Principal component 2 and total score have a correlation of -0.06927. Given the interpretation of Principal component 1 and 2 in part d, the total score appears to be assessing the overall athletic ability of the decathlete (Factor 1) and not their upper body strength (Factor 2).