Homework 2 (Solution)

1. Since the p-values for the testing parallel, coincidental and horizontal are < 0.0001, 0.0007 and < 0.0001 respectively. So we reject all the three null hypothesis and conclude that the two profiles are not parallel, coincidental nor horizontal.

2. (a) The principal components analysis based on S shows that only linear combination of the traits can capture 99.42% of the total variance. While the result based on R shows that 5 linear combinations can capture 95.31% of the total variance. But here I think it is more appropriate to use principal components analysis based on R because R is invariant to scale transformation.
(b) After transferring units for $y_6$ from degree to radians, the values for $y_6$ decreased and the eigenvalues in S changed a lot while R remains the same. The total variance is much smaller.
(c) It depends on the minimum percent total variation you set. You can retain the first 3 or 4 components. If you retain the first 3, then it will capture 75.43% of the total variance. If you retain the first 4, then it will capture 87.09% of the total variance.

3. (a) By using the PC analysis, we can get the first two eigenvalues of R are greater than 1. Two common factors will account for a cumulative proportion of 87.14% of the total sample variance. And since all the variables load highly on the first factor, we can view the first factor as overall affordable factor while the second factor can be interpreted as basic household necessity factor. So we retain the first two.
(b) The result by using the principal factors method also retains two factors, similar as the result in part (a). The total communalities estimated by the principal factors method is 4.655 a little bit lower than 5.228 (the result by the principal components method).
(c) The p-value for testing that one factor is sufficient is very small, while the p-value for testing that two factors are sufficient is around 0.49. So two factors are sufficient.
(d) The more items a person bought in categories such as cloth, home furnishings, etc. the higher the estimated values of factor 1 are. The estimated values of factor 2 positively correlated with the numbers of items bought in electronic and sporting goods.