# Hotel Tourism - Solution:

The part of the questionnaire made available is just on what is important when choosing vacation spot, and contains nothing specific to the hotels. Without more information we are therefore unable to answer the last question. We may be able to throw some light on the following:

What are the factors underlying the choice of vacation spot?
How can satisfaction be explained?
What separates the areas wrt. guest profile?
Which groups/segments exist within each area?

To get an understanding on the underlying decisive factors for choice of vacation spot, we may start by performing a factor analysis on the correlation matrix of the coded 19 responses to the given question from the questionnaire. These responses are ordinal on a 9-point scale, and not measured on a continuous (interval) scale required by the theory. In practice people are seldom worried about this. The uncovering of differences between the three locations may be based on the factor analysis or alternatively a discriminant analysis.

A so-called Scree plot of eigenvalues shows that 3 of them are well above one, and two slightly above one. This means that we will consider five underlying factors, possible reduced to three if no reasonable interpretation of the last two may be given. In the case of factor analysis on the covariance (or equivalently on the standardized variables) an eigenvalue less than one means a factor less important than a single variable.

.

Scree Plot of Nature; ...; Clean nature
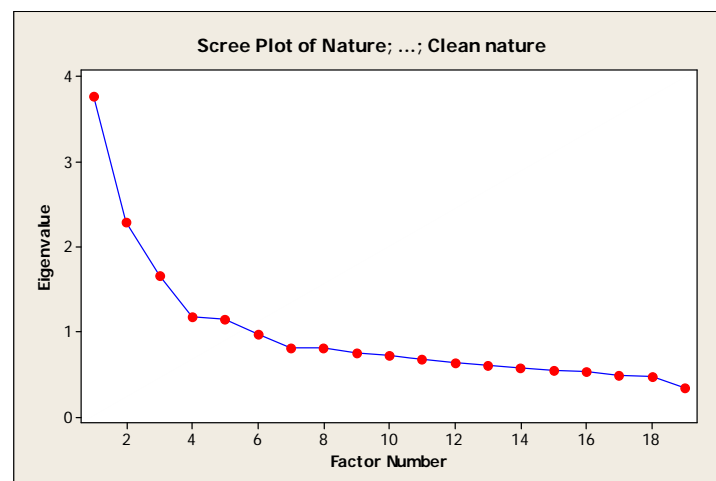
Below is the result from an extraction of five factors from the correlation matrix of the 19 variables, using principal component extraction and varimax rotation. For brevity we omit the details of the unrotated solution. We see that the 5 factors explain 52.9% of the (co)variation in the 19 variables, which is more evenly distributed among the 5 factors in the rotated solution than the unrotated one.

Factor Analysis: Nature; Bathing; Hiking; Entertainmen; Peace; Culture; Nice We

```
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities
        570 cases used  199 cases contain missing values

…… omitted for brevity

Variance         3.7742   2.2995   1.6589   1.1764   1.1441      10.0531
% Var            0.199    0.121    0.087    0.062    0.060        0.529
```

```
Rotated Factor Loadings and Communalities
Varimax Rotation

Variable          Factor1  Factor2  Factor3  Factor4  Factor5  Communality
Nature            0.293   -0.591   -0.058   -0.141    0.049       0.461
Bathing           0.001    0.088   -0.130    0.747   -0.029       0.583
Hiking            0.089   -0.320   -0.036    0.502    0.513       0.627
Entertainment     0.072    0.516    0.300    0.275   -0.269       0.509
Peace            -0.175   -0.736    0.113   -0.065   -0.204       0.631
Culture           0.227   -0.232    0.570   -0.303    0.175       0.552
Nice Weather      0.254   -0.154    0.024    0.136   -0.690       0.584
New Exper         0.812   -0.015    0.031   -0.024   -0.234       0.716
Children friendly -0.178    0.183    0.072    0.667   -0.120       0.530
Good Food         0.200   -0.063    0.124    0.055   -0.673       0.515
Excursions        0.548   -0.169    0.260   -0.046   -0.159       0.424
Cheap             0.168   -0.275    0.369   -0.007   -0.381       0.386
Uncommon          0.739    0.018    0.188   -0.018   -0.169       0.610
Community         0.217    0.054    0.317    0.405   -0.203       0.356
Know area         0.497   -0.437    0.133    0.108    0.070       0.472
Get in shape      0.074   -0.166    0.448    0.481    0.134       0.483
Shopping         -0.016    0.110    0.692    0.094   -0.328       0.608
Meet people       0.266   -0.020    0.634    0.058   -0.081       0.483
Clean nature      0.171   -0.650    0.162    0.060   -0.198       0.520

Variance         2.2520   2.1440   1.9363   1.8951   1.8257      10.0531
% Var            0.119    0.113    0.102    0.100    0.096        0.529

Factor Score Coefficients

Variable          Factor1  Factor2  Factor3  Factor4  Factor5
Nature            0.107   -0.258   -0.103   -0.027    0.037
Bathing           0.044   -0.013   -0.164    0.419   -0.026
Hiking            0.100   -0.165   -0.024    0.310    0.339
Entertainment     0.024    0.264    0.145    0.085   -0.089
Peace            -0.269   -0.412    0.046   -0.002   -0.182
Culture           0.018   -0.030    0.391   -0.203    0.229
Nice Weather      0.021   -0.086   -0.167    0.078   -0.421
New Exper         0.447    0.103   -0.170   -0.007   -0.025
Children friendly -0.105    0.022    0.006    0.345   -0.078
Good Food        -0.016   -0.035   -0.072    0.017   -0.398
Excursions        0.232    0.001    0.042   -0.035    0.013
Cheap            -0.069   -0.120    0.144   -0.024   -0.181
Uncommon          0.388    0.121   -0.038   -0.022    0.034
Community         0.053    0.039    0.103    0.189   -0.042
Know area         0.224   -0.153   -0.026    0.084    0.125
Get in shape     -0.038   -0.080    0.256    0.235    0.171
Shopping         -0.175    0.067    0.417   -0.029   -0.100
Meet people       0.025    0.053    0.361   -0.028    0.091
Clean nature     -0.053   -0.321    0.006    0.061   -0.11
```

The each factor of the rotated solution we may look at the size (in absolute value) of its factor loadings, in order see what characterizes the variables that load on this factor. From this we are tempted to give the five factors the following interpretation:

Factor 1: Adventures (Uncommon, New experience)
Factor 2: Environment (Peace, Cleanliness, Nature) vs Entertainment
Factor 3: Town activity (Shopping, Meet people, Culture)
Factor 4: Physical activity (Bathing, Hiking, Get in shape, Children friendly)
Factor 5: Passive enjoyment (Good Food, Nice weather) vs Active (Hiking)

Factor 2 and 5 are so-called bipolar factor with both large positive and negative coefficients that separates two categories of people measured along a dimension.

While the factor loadings express how each variable is determined linearly by the five unobserved underlying (latent) factors, the factor scores express how each (unobserved) factor is determined as a linear combination of the 19 observed variables. This may be computed for all respondents having answered all 19 items. The result may be presented graphically in several different manners. One is by a collection of box-plots, displaying the median, quartile range, the enlarged range and the outlying scores. We see that Region 1 (Valleyland) has the highest median score on Factor 3 (Town activity), Region 2 (Fjordland) highest on Factor 5 (Passive enjoyment vs, hiking) and Region 3 (Coastland) highest on Factor 1 and 4 (Adventures and Physical activity). For Factor 2 (Environment vs Entertainment) the medians in the three regions are about equal. This may make sense to those who know the actual regions, which are Voss, Stryn and Vesterålen respectively. From the plot we may also see the regions with lowest median score and the ones with small and large variation in scores and possibly give sensible interpretations.

Boxplot of Factorscores for 5 factors for each of 3 regions
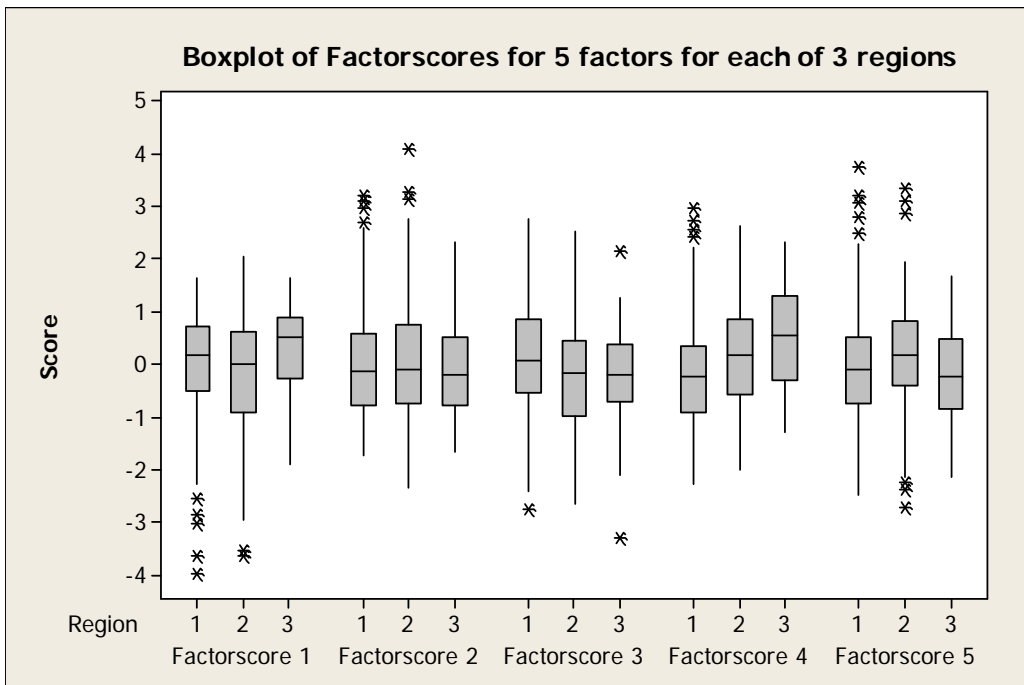
We may alternatively perform a discriminant analysis, where the observed 19 variables are used as predictors of Region. Here is the output, where the coefficients for three linear predictor formulas are listed at the end. For a respondent, take the 19 coded answers, weigh them by the coefficients and sum. The respondent is then "predicted" to belong to the region with highest score. The importance of the various variables in the prediction may be judged from the coefficients. For instance Culture is given more weight in being classified to Region 1 than the other regions. The output also contains information on how successful the procedure was to "predict" the actual observation, as an indication on a lower limit to success in predicting new observations. An alternative mode of discriminant analysis exists, more directed towards response profiles.

Discriminant Analysis: Region versus Nature; Bathing; ...

```
Linear Method for Response: Region


Predictors: Nature; Bathing; Hiking; Entertainment; Peace; Culture;
            Nice Weather; New Exper; Children friendly; Good Food; Excursions;
            Cheap; Uncommon; Community; Know area; Get in shape; Shopping;
            Meet people; Clean nature


Group          1         2         3
Count        346       166        58

570 cases used, 199 cases contain missing values

Summary of classification

                     True Group
Put into Group      1        2        3
1                 208       52       14
2                  82       69        9
3                  56       45       35
Total N           346      166       58
N correct         208       69       35
Proportion      0.601    0.416    0.603

N = 570            N Correct = 312          Proportion Correct = 0.547

Squared Distance Between Groups

         1        2        3
1  0.00000  0.71126  1.40987
2  0.71126  0.00000  1.03271
3  1.40987  1.03271  0.00000


Linear Discriminant Function for Groups

                         1         2         3
Constant            -43.071   -41.179   -47.197
Nature                4.843     4.944     5.275
Bathing               0.904     0.965     0.975
Hiking                0.282     0.467     0.378
Entertainment         1.158     1.121     0.992
Peace                 0.669     0.648     0.456
Culture               0.409     0.296     0.162
Nice Weather          0.741     0.626     0.884
New Exper             1.947     1.789     1.872
Children friendly     0.694     0.899     1.163
Good Food             0.533     0.652     0.713
Excursions            0.375     0.453     0.383
Cheap                 0.584     0.520     0.448
Uncommon             -0.358    -0.427    -0.171
Community             0.104     0.064     0.056
Know area             0.815     0.745     0.933
Get in shape         -0.059    -0.133    -0.049
Shopping              0.004    -0.117    -0.099
Meet people           0.098     0.146     0.141
Clean nature          2.105     1.962     2.133
```