**Principal components analysis**

**QUESTION** **- 2003**

Some behavioural ecologists wished to establish the feeding preferences of a desert rodent. They contrasted five seeds that were collected from within the rodent’s burrow with five seeds that had been collected randomly from the surrounding habitat. They collected the following data from each seed: protein (% of dry mass), carbohydrate (% of dry mass), lipid content (% of dry mass), moisture content (%), salt content (mg/g), mass (g) and length (mm).

1. What is a principal component? (4 marks)
2. How much variance is explained by the first two principal components? (2 marks)
3. Does the ordination provide a test of the null hypothesis that the rodents are selecting seeds at random from the surrounding habitat? (6 marks)
4. What attributes of the seeds does the ordination suggest are most important to the rodent? (8 marks)

**QUESTION** **5: 2009**

A marine ecologist was studying what attributes of coral reefs were associated with territories of damselfish. From five known territories in the field, she measured six environmental variables: 1) depth (m), 2) coral cover (%), 3) algal cover (%), 4) sediment load (g), 5) average wave heights (m) and 6) reef slope (). These five sites were then compared to five sites randomly positioned on the reef that lacked territories.

The following is the output from a principal components analysis of these variables:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Eigenvalue | 3.8018 | 1.0703 | 0.9981 | 0.0959 | 0.0315 | 0.0024 |
| Proportion | 0.634 | 0.178 | 0.166 | 0.016 | 0.005 | 0.000 |
| Cumulative | 0.634 | 0.812 | 0.978 | 0.994 | 1.000 | 1.000 |
|  |  |  |  |  |  |  |
| Variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
| Depth | -0.501 | -0.085 | 0.131 | 0.215 | -0.706 | 0.424 |
| Coral cover | -0.485 | -0.172 | -0.079 | -0.850 | 0.073 | -0.031 |
| Algal cover | 0.030 | -0.811 | -0.538 | 0.201 | 0.081 | 0.073 |
| Sediment | 0.049 | -0.537 | 0.827 | 0.018 | 0.136 | -0.087 |
| Wave height | -0.509 | 0.015 | -0.046 | 0.318 | -0.031 | -0.798 |
| Slope | 0.501 | -0.134 | -0.041 | -0.299 | -0.685 | -0.412 |

1. Does a two-dimensional ordination successfully summarise these data? State reasons for your answer. (3 marks)
2. Which variable is most strongly correlated with the second principal component? (2 marks)
3. Does the second principal component help in distinguishing between the sites with and without territories? (5 marks)
4. What characteristics of the reef sites are most closely associated with the presence of damselfish territories? State the reasons for your decision. (10 marks)

**QUESTION 3.** **2010**

An agricultural scientist was trying to find out what soil properties were associated with disease in a potato crop. He collected ten soil samples from sites where the potatoes were diseased and ten soil samples from sites where the crop was free of disease. From each sample, he measured the following data: % organic matter, thickness of leaf litter above the soil, surface moisture content, % sand and % clay. The following is a principal components analysis and score plot of the data set.

Principal Component Analysis: % organic matter, leaf litter, moisture, % sand, % clay

Eigenvalue 3.1309 1.1635 0.4673 0.2374 0.0009

Proportion 0.626 0.233 0.093 0.047 0.000

Cumulative 0.626 0.859 0.952 1.000 1.000

Variable PC1 PC2 PC3 PC4 PC5

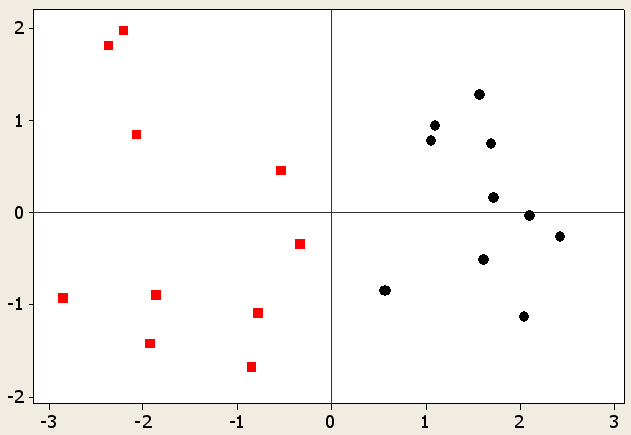
% organic matter -0.548 0.116 -0.133 -0.393 0.717

Leaf litter -0.546 0.112 -0.137 -0.431 -0.697

Moisture -0.386 -0.448 0.797 0.127 -0.005

% sand 0.503 -0.073 0.329 -0.795 0.021

% clay 0.013 -0.876 -0.470 -0.105 0.007



**First component**

**Second component**

Diseased

Healthy

Answer all parts 3A-D.

**3A)** Does the two-dimensional ordination successfully summarise the multivariate data? Give reasons for your answer. (4 marks)

**3B)** What soil measurements are most closely associated with the presence of disease? State the reasons for your decision. (8 marks)

**3C)** Some farmers believed the disease was due to a build up of fungi in the leaf litter. Does the analysis support this hypothesis? (2 marks)

**3D)** Briefly describe an experiment that could test the farmers’ hypothesis (6 marks).

**QUESTION 3. 2011**

A forest biologist was studying diet selection by wallabies to understand what properties of seedlings were associated with higher rates of herbivory following plantings. She ran some field trials with seedlings and collected five seedlings which had been damaged and five which were left alone. The seedlings were then returned to the laboratory and several plant quality traits were measured from each seedling. A principal components analysis was conducted with the following variables: moisture content (%), carbohydrate (% of dry mass), lipid content (% of dry mass), leaf length (cm), toughness, concentration of terpenes (a chemical defence), and protein (% of dry mass). The following is the output from the PCA:

Seedling

damaged

undamaged

3

2

1

0

-

1

-

2

2

.

5

2

.

0

1

.

5

1

.

0

-

0

.

5

-

1

.

0

-

Principal Component Analysis: moisture, carbohydrate, lipid, leaf length, toughness, terpenes, protein

Eigenanalysis of the Correlation Matrix

Eigenvalue 2.2427 1.8376 1.1920 0.9568 0.3977 0.3473 0.0259

Proportion 0.320 0.263 0.170 0.137 0.057 0.050 0.004

Cumulative 0.320 0.583 0.753 0.890 0.947 0.996 1.000

Variable PC1 PC2 PC3 PC4 PC5 PC6 PC7

moisture -0.509 -0.061 0.031 -0.529 0.443 0.430 -0.275

carbohydrate 0.543 -0.287 0.072 -0.398 -0.104 -0.195 -0.640

lipid 0.047 0.340 0.694 -0.338 -0.452 0.220 0.183

leaf length -0.416 -0.162 0.566 0.204 0.180 -0.610 -0.186

toughness -0.502 0.132 -0.396 -0.159 -0.657 -0.229 -0.255

terpenes -0.057 -0.657 -0.050 -0.423 -0.116 -0.186 0.580

protein 0.124 0.569 -0.180 -0.449 0.334 -0.519 0.213

Answer all parts 3A to 3C.

**3A)** Brieflydescribe what principal components analysis is aiming to do with these data? (5 marks)

**3B)** What traits of the seedlings are most closely associated with being damaged by wallabies? State the reasons for your decision. (10 marks)

**3C)** “The score plot provides evidence to reject the null hypothesis that the damaged seedlings were of equal qualities to the undamaged seedlings” Is this statement true or false? Give reasons for your decision. (5 marks).

**QUESTION 2.**  2012

To understand how food quality might influence the population dynamics of plague locusts, an entomologist ran feeding trials in the laboratory with artificial diets varying in nutritional value. Each of five locusts was offered a choice of two diets, and when they started feeding, the diet that they had selected was removed and analysed for seven nutritional variables (carbohydrate, lipid content, salt, mass, moisture, protein, vitamins). The scientist then measured the same properties from the rejected diets. A principal components analysis was conducted to contrast the selected and rejected diets with the following output:

Principal Component Analysis: protein, carbohydrate, salt, moisture, lipid, vitamins

Eigenanalysis of the Covariance Matrix

Eigenvalue 40.164 10.989 7.217 3.919 2.207 0.632 0.028

Proportion 0.616 0.169 0.111 0.060 0.034 0.010 0.000

Cumulative 0.616 0.785 0.896 0.956 0.990 1.000 1.000

Variable PC1 PC2 PC3 PC4 PC5 PC6 PC7

carbohydrate 0.660 0.601 0.390 0.043 -0.144 0.132 -0.109

lipid 0.044 0.101 0.017 0.009 0.475 -0.675 -0.553

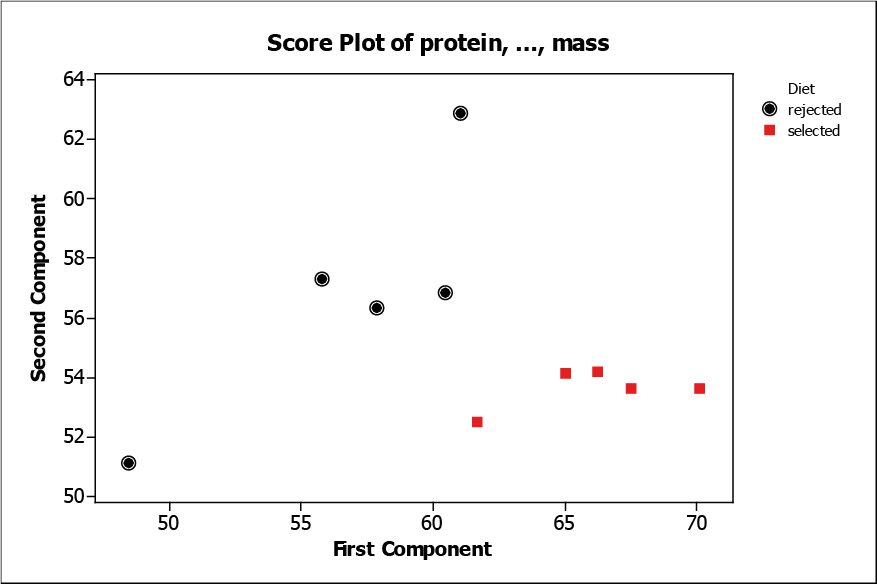
mass -0.103 0.611 -0.749 -0.144 -0.139 -0.096 0.074

moisture 0.012 0.210 0.067 0.285 0.650 -0.051 0.667

protein 0.644 -0.391 -0.260 -0.455 0.014 -0.264 0.298

salt 0.370 -0.241 -0.453 0.682 0.093 0.255 -0.248

vitamins -0.015 -0.027 0.099 0.474 -0.550 -0.617 0.285



Answer all parts 2A to 2C.

**2A)** Does the score plot effectively summarise the multivariate data set? State reasons for your answer (5 marks)

**2B)** What properties of the diets are most closely associated with being selected by the locusts? State the reasons for your decision. (10 marks)

**2C)** A colleague suggested running seven two-sample t-tests (one for each variable) that contrasted the five selected and five rejected diets. With reasons, discuss whether you think this is a good idea or not. (5 marks)

**QUESTION 3.**  2013

A group of farmers is concerned about the possibility of air and water contamination near wells used to extract coal seam gas on agricultural land. They contract an environmental chemist to measure the concentrations of methane and some of the volatile organic compounds of environmental concern (toluene, xylene, benzene, and ethyl benzene). The chemist sampled ground water directly adjacent to five wells, and also five control sites from the same area that are at least one km from a well site. A principal components analysis was conducted to contrast the well sites to the control sites with the following output:

Principal Component Analysis: methane, toluene, xylene, benzene, ethyl benzene

Eigenanalysis of the Correlation Matrix

Eigenvalue 2.2921 1.2872 0.8909 0.4154 0.1144

Proportion 0.458 0.257 0.178 0.083 0.023

Cumulative 0.458 0.716 0.894 0.977 1.000

Variable PC1 PC2

methane 0.446 -0.434

toluene 0.592 -0.026

xylene 0.623 0.152

benzene 0.228 0.725

ethyl benzene 0.099 -0.512



Answer all parts 3A to 3C.

**3A)** Write a paragraph to explain to the farmers whether the score plot does or does not suggest a problem with groundwater contamination. (5 marks)

**3B)** Which contaminants, if any, are most closely associated with proximity to the well sites? State the reasons for your decision. (10 marks)

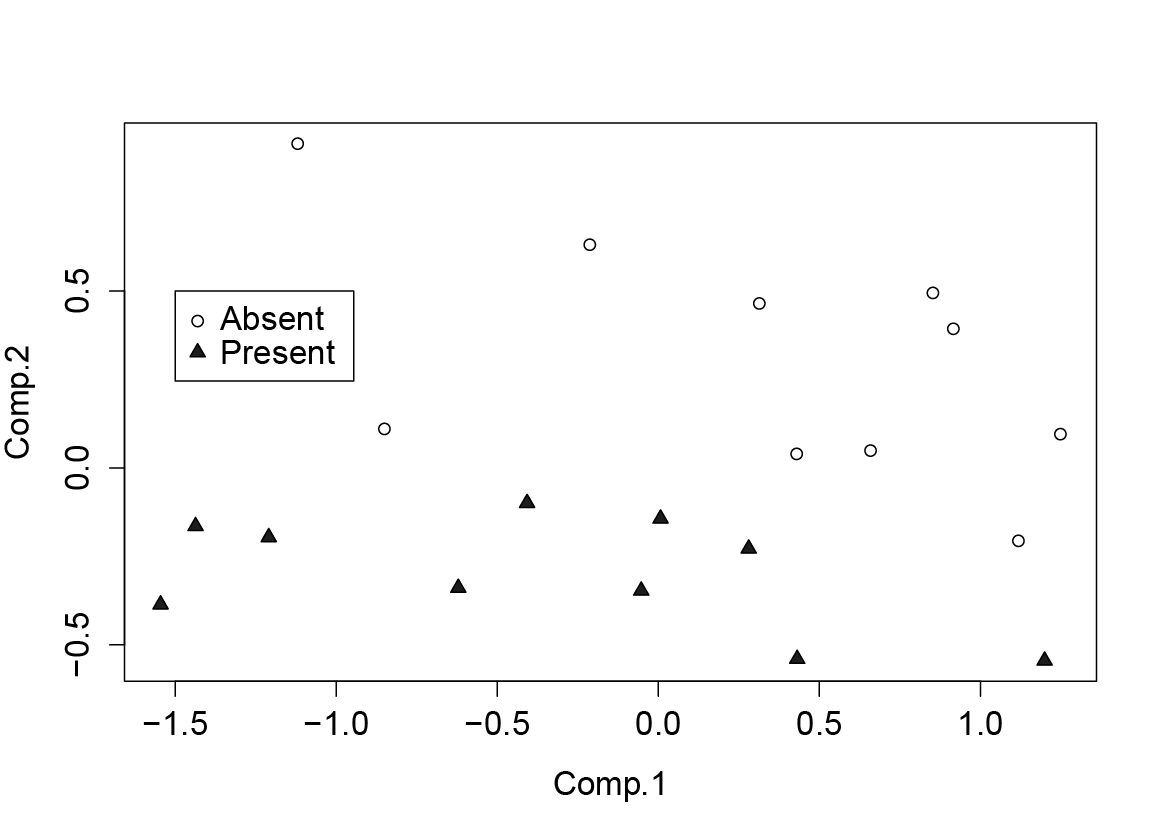
**3C)** In her final report, the environmental chemist stated “The coal seam gas wells are clearly responsible for groundwater contamination” Is she justified in this conclusion? (5 marks)

**QUESTION 3.**  2014

A research student is trying to understand why seagrass beds in urban estuaries have declined so dramatically over the past few decades. She measured five abiotic variables (mean salinity, mean temperature, depth, sediment grain size and nitrogen concentration) in ten areas where seagrass beds persist, and in ten areas where they are now absent but were present historically. She conducted a principal components analysis and obtained the following output:

|  |
| --- |
| Importance of components: |
| Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 |
| Standard deviation 0.8741857 0.3900258 0.29577863 0.17472485 0.100759900 |
| Proportion of Variance 0.7316516 0.1456410 0.08375881 0.02922848 0.009720137 |
| Cumulative Proportion 0.7316516 0.8772926 0.96105138 0.99027986 1.000000000 |

|  |
| --- |
| Loadings: |
| Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 |
| Salinity 0.152 0.914 0.351 0.123 |
| Temperature -0.957 0.252 0.141 |
| Grain\_size -0.146 0.125 -0.980 |
| Depth -0.226 -0.797 -0.135 0.527 0.133 |
| Nitrogen 0.526 -0.333 0.773 |



Answer all parts 3A to 3B.

**3A)** Which abiotic variables, if any, are most closely associated with the loss of seagrass beds? State the reasons for your decision. (10 marks)

**3B)** Several local councils are attempting to manage stormwater input into the estuaries with the aim of restoring seagrass beds. Discuss whether the principal components analysis supports the idea that better management of stormwater would benefit seagrass beds (10 marks)

**QUESTION???. 2004**

A behavioural ecologist was studying what attributes of plumage affect mating success in Rainbow Lorikeets (*Trichoglossus haematodus*). These parrots are very brightly coloured with a red beak, a red eye, a blue head, an orange/yellow breast, and green wings and tail. From an aviary population, she measured six traits from five males that had successfully mated, and the same traits from five males that had been unsuccessful in the current breeding season. The traits measured were; 1) intensity of beak colour, 2) intensity of eye colour, 3) intensity of blue on the head, 4) intensity of green on the wings, 5) area of orange on the breast, and 6) area of yellow on the breast.

The following is the output from a principal components analysis of these traits:

Not mated

Mated

-

3

-

2

-

1

0

1

2

3

-

2

-

1

0

1

2

PC1

PC2

Eigenvalue 3.8018 1.0703 0.9981 0.0959 0.0315 0.0024

Proportion 0.634 0.178 0.166 0.016 0.005 0.000

Cumulative 0.634 0.812 0.978 0.994 1.000 1.000

Variable PC1 PC2 PC3 PC4 PC5 PC6

Beak -0.501 -0.085 0.131 0.215 -0.706 0.424

Eye -0.485 -0.172 -0.079 -0.850 0.073 -0.031

Blue 0.030 -0.811 -0.538 0.201 0.081 0.073

Green 0.049 -0.537 0.827 0.018 0.136 -0.087

Orange -0.509 0.015 -0.046 0.318 -0.031 -0.798

Yellow 0.501 -0.134 -0.041 -0.299 -0.685 -0.412

1. Which variable is most strongly correlated with the first principal component? (2 marks)
2. Does the second principal component help in distinguishing between the successful and unsuccessful males? (2 marks)
3. What characteristics of the birds’ plumage are most closely associated with mating success? State the reasons for your decision. (8 marks)
4. With reasons, discuss whether principal components analysis (PCA), multidimensional scaling (MDS) is the best analysis for addressing the researcher’s aims? Can you suggest any other analyses that might have been used? (8 marks)

**QUESTION ?? 2005**

During an outbreak of plague locusts in western NSW, agricultural researchers were attempting to understand which plant traits made a crop species more susceptible to attack. From ten fields in which locusts were present, and ten fields in which they were absent, they obtained measures of: 1) average plant height, 2) average leaf number per plant, 3) nitrogen content of plant tissues, 4) tannin levels of plant tissues, and 5) soil moisture.

The following is the output from a principal components analysis of these plant traits:

Principal Component Analysis: Plant height, Leaf number, N content, Tannin level, Soil moisture

Eigenvalue 3.1309 1.1635 0.4673 0.2374 0.0009

Proportion 0.626 0.233 0.093 0.047 0.000

Cumulative 0.626 0.859 0.952 1.000 1.000

Variable PC1 PC2 PC3 PC4 PC5

Plant height -0.548 0.116 -0.133 -0.393 0.717

Leaf number -0.546 0.112 -0.137 -0.431 -0.697

N content -0.386 -0.448 0.797 0.127 -0.005

Tannin levels 0.503 -0.073 0.329 -0.795 0.021

Soil moisture 0.013 -0.876 -0.470 -0.105 0.007



1. Does the two-dimensional ordination successfully summarise the multivariate data? Give reasons for your answer (4 marks)
2. What plant traits are most closely associated with the presence of locusts? State the reasons for your decision. (8 marks)
3. Some farmers blamed the lack of available water for irrigation on the locust plague. What does the analysis suggest about the importance of available water? (4 marks)
4. “The score plot provides evidence to reject the null hypothesis that plant traits were equal between the fields with and without locusts” True or False? Give reasons for your decision. (4 marks).

**QUESTION 6: 2006**

An environmental scientist was trying to improve the attractiveness of baits used to control foxes in a national park. Baits were placed in the field that differed in nutritional content, mass and the depth to which they were buried. At the end of the trial, the bait types taken by foxes were compared to those that were left uneaten. A principal components analysis was conducted with the following variables: protein (% of dry mass), carbohydrate (% of dry mass), lipid content (% of dry mass), moisture content (%), salt content (mg/g), mass (g) and depth buried (cm).

The following is the output from the PCA:

Principal Component Analysis: protein, carbohydrate, lipid, moisture, salt, mass, depth

Eigenanalysis of the Correlation Matrix

Eigenvalue 2.2427 1.8376 1.1920 0.9568 0.3977 0.3473 0.0259

Proportion 0.320 0.263 0.170 0.137 0.057 0.050 0.004

Cumulative 0.320 0.583 0.753 0.890 0.947 0.996 1.000

Variable PC1 PC2 PC3 PC4 PC5 PC6 PC7

protein -0.509 -0.061 0.031 -0.529 0.443 0.430 -0.275

carbohydrate 0.543 -0.287 0.072 -0.398 -0.104 -0.195 -0.640

lipid 0.047 0.340 0.694 -0.338 -0.452 0.220 0.183

moisture -0.416 -0.162 0.566 0.204 0.180 -0.610 -0.186

salt -0.502 0.132 -0.396 -0.159 -0.657 -0.229 -0.255

mass -0.057 -0.657 -0.050 -0.423 -0.116 -0.186 0.580

depth 0.124 0.569 -0.180 -0.449 0.334 -0.519 0.213



1. Briefly describe what the first principal component (PC1) represents? 4 marks
2. What attributes of the baits are most closely associated with being eaten by foxes? State the reasons for your decision. 8 marks
3. “The score plot provides evidence to reject the null hypothesis that the attributes of baits were equal between those eaten and those uneaten” Is this statement true or false? Give reasons for your decision. 4 marks.
4. Many people argue that the baits should be buried deeply to avoid ingestion by native species. How does this practice affect the likelihood of baits being consumed?

4 marks

**QUESTION 6:2007**

An environmental scientist was studying what attributes of floodplains are associated with the dieback of river red gums. He surveyed 10 sites, half of which had extensive dieback, and half of which had healthy trees. He measured six variables from each of the sites: 1) years since last flood, 2) soil salinity, 3) abundance of herbivorous insects, 4) density of trees, 5) distance from river bed, and 6) volume of last flood. The following is the output from a principal components analysis of these variables:

2

PC2

1

Healthy

With dieback

0

1

-

2

-

3

2

1

0

1

-

2

-

3

-

PC1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Eigenvalue | 3.8018 | 1.0703 | 0.9981 | 0.0959 | 0.0315 | 0.0024 |
| Proportion | 0.634 | 0.178 | 0.166 | 0.016 | 0.005 | 0.000 |
| Cumulative | 0.634 | 0.812 | 0.978 | 0.994 | 1.000 | 1.000 |
|  |  |  |  |  |  |  |
| Variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
| Years since last flood | -0.501 | -0.085 | 0.131 | 0.215 | -0.706 | 0.424 |
| Soil salinity | -0.485 | -0.172 | -0.079 | -0.850 | 0.073 | -0.031 |
| Abundance of herbivorous insects | 0.030 | -0.811 | -0.538 | 0.201 | 0.081 | 0.073 |
| Density of trees | 0.049 | -0.537 | 0.827 | 0.018 | 0.136 | -0.087 |
| Distance from river bed | -0.509 | 0.015 | -0.046 | 0.318 | -0.031 | -0.798 |

Volume of last flood 0.501 -0.134 -0.041 -0.299 -0.685 -0.412

a) Which variable is most strongly correlated with the first principal component? (2 marks)

b) Does the second principal component help in distinguishing between the sites with and without dieback? (2 marks)

c) What habitat variables are most closely associated with dieback of river red gums? State the reasons for your decision. (8 marks)

d) With reasons, state why a principal components analysis (PCA) is a useful approach for examining this data set. Can you suggest any other analyses that might have been used? (8 marks)

**QUESTION??:** 2008

Some biologists were studying the feeding preferences of mangrove crabs. They contrasted five leaves that had been taken into crab burrows for consumption with five leaves that had been avoided by the crabs. They collected the following data from each leaf: mass (g) and length (mm), C:N ratio, concentration of tannins (% of dry mass), lipid content (% of dry mass), moisture content (%) and salt content (mg/g).

The following is the output from a principal components analysis:

-2

-1

0

1

2

3

PC1

PC2

0

1

2

-1

-2

avoided

consumed

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
| Eigenvalue | 3.4908 | 2.1710 | 0.6688 | 0.4459 | 0.1859 | 0.0236 | 0.0141 |
| Proportion | 0.499 | 0.310 | 0.096 | 0.064 | 0.027 | 0.003 | 0.002 |
| Cumulative | 0.499 | 0.809 | 0.904 | 0.968 | 0.995 | 0.998 | 1.000 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
| salt | 0.001 | -0.540 | 0.649 | 0.355 | -0.393 | 0.048 | 0.062 |
| tannins | 0.167 | 0.542 | 0.286 | 0.660 | 0.285 | -0.278 | -0.049 |
| length | -0.375 | 0.399 | -0.232 | 0.206 | -0.763 | 0.044 | 0.141 |
| C:N ratio | 0.178 | -0.469 | -0.649 | 0.536 | 0.020 | -0.165 | -0.107 |
| mass | -0.518 | -0.134 | 0.027 | -0.118 | 0.057 | -0.804 | 0.222 |
| lipid | -0.516 | -0.010 | -0.093 | 0.300 | 0.277 | 0.468 | 0.583 |
| moisture | -0.515 | -0.134 | 0.106 | 0.068 | 0.319 | 0.162 | -0.758 |

1. What do the symbols on the plot represent? What does the distance between the symbols represent? (5 marks)
2. What do the numbers in the columns of the lower table represent? (5 marks)
3. Does the ordination provide a test of the null hypothesis that the crabs show no selection of leaves based on their nutritional properties? (2 marks)
4. What nutritional properties of the leaves does the ordination suggest are most important to the crabs? (8 marks)