SaRD Full Practice Quiz 1

120 minute time limit.

Open book, but no collaboration – so no facebook, email, instant messaging etc. etc., but notes, internet, books, your scripts etc. are all fine.

Round answer to appropriate numbers of decimals places.

The quiz comprises five sections, each of which will be worth 10 marks/20% of the total quiz.

Answer each question in a clearly annotated script. You will submit the script for marking.

Make sure your name and date is at the top of your quiz script. You should use the format of scripts like the one that was given to you in the first week (see R-script for P1 exercises).

In questions where you are asked to analyse data, you must follow the analysis pathways set out in the lectures, e.g.:

1. Write clear hypotheses
2. Explore descriptive statistics – means, standard deviation, distribution etc
3. Create appropriate plots to visualise/explore data
4. Decide on test(s) to be done (see flow chart)
5. Examine assumptions of tests: normality, variances, independence
6. Carry out tests
7. Report your results: results statements / plots

Section 1

The dataset *InsectSprays* contains data on insect density in areas treated with different kinds of pesticide spray. Load the data into R. Investigate if the different kinds of spray are effective in reducing insect density – i.e. is there a difference in insect density between sprays.

Marking guidance

1 mark for the hypothesis; 2 marks for preliminary data exploration and visualisation, 3 marks for testing the assumptions and reporting, 2 marks for selection and carrying out correct test, 2 marks reporting your results and presenting the data.

Section 2

The dataset *Raccoon* has three variables: weight (lb), sex (male and female) and population location (N, S, E, W) for a number of animals. Load the dataset into R and investigate how weight differs with sex and/or population location.

Marking guidance

1 mark for the hypothesis; 1 marks for preliminary data exploration and visualisation, 3 marks for testing the assumptions and reporting, 3 marks for selection and carrying out correct test, 2 marks reporting your results and presenting the data.

Section 3

Part 1

The figures below are taken from Garbutt & Wolters (2008) "The natural regeneration of salt marsh on formerly reclaimed land" published in *Applied Vegetation Science.*
Note: de-embankment means the point at which regeneration began. Reference marshes are natural marshes.

a) For each figure write a sentence, suitable for a results section, that describes the contents of the figure. [2 marks]

b) Using Fig 2 (left-hand plot), calculate the difference in species richness if years since de-embankment is 136 (give your answer as an integer). [1 mark]

c) Briefly state the confidence you would place in the relationships in the figures below and why you think that this is the case. [2 marks]

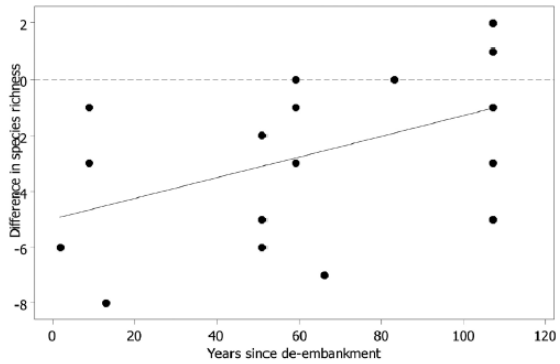

Fig. 2. Difference *ds* in species richness between de-embankment and reference marshes over time: $ds = -5.01 + 0.037a$, where a = years since de-embankment. $R^2 = 21.9\%$; $p = 0.05$. Negative values indicate fewer species within a de-embankment site than adjacent reference marsh.

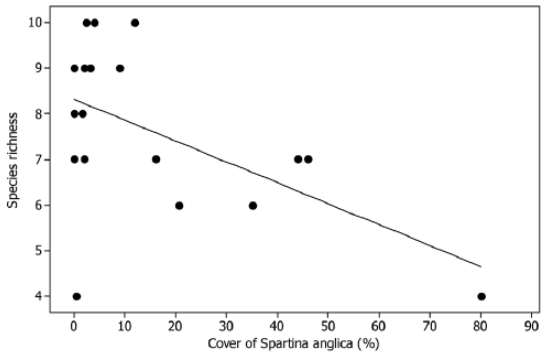

Fig. 3. Total species richness sr in relation to mean percentage cover of *Spartina anglica* within each de-embankment site: $sr = 8.3 - 0.04560a$, where a = % *S. anglica*; R^2=30.3 %; $p = 0.018$.

Part 2

Below is the output from R of a model investigating the effects of climate on the net primary productivity of a series of temperate forests.

The variables are:
NPP – net primary productivity
MAT – mean annual temperature
MAP – mean annual precipitation
DS – an indicator of drought stress
TCM – temperature of the coldest month

```
mod<-lm(NPP~MAT+DS+MAP+TCM,data=dat)
> summary(mod)

Call:
lm(formula = NPP ~  MAT +  DS + MAP +TCM, data = dat)

Residuals:
   Min      1Q      Median      3Q     Max
-4.7139 -1.3019 -0.0719       1.1764  6.9493

Coefficients:
                    Estimate      Std. Error    t value       Pr(>|t|)
(Intercept)         1.858e+01     3.341e+00     5.559         7.91e-08 ***
MAT                 3.896e-01     1.210e-01     -3.220        0.001481 **
Drought_stress      5.489e-03     1.436e-03     -3.823        0.000172 ***
MAP                 -6.154e-04    3.061e-04     -2.010        0.045631 *
TCM                 -5.2457e-01   5.321e+00     -1.010        0.078542 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.085 on 217 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.1993,    Adjusted R-squared:  0.1698
F-statistic: 6.751 on 8 and 217 DF,  p-value: 6.707e-08
```

a) Write a sentence suitable for a results section that describes the findings of this model. [3 marks]

b) What would you do next if you got this output? Describe the procedures you would carry out next [2 marks]

Section 4

The dataset *bodyfat* contains data on the percentage body fat from adult human with a range of other body measurements and age. Load the data into R. Investigate which of the variables below are good predictors (explanatory variables) of percentage body fat.

Explanatory variables: age, weight, height, chest, hip, ankle

Marking guidance

1 mark for the hypothesis; 1 mark for preliminary data exploration and visualisation; 2 marks for testing and reporting collinearity, 4 marks for model building and validation, 2 marks reporting your results.

Section 5

The dataset *Titanic* contains data on the frequency of different passenger categories that either survived or perished during the Titanic passenger liner tragedy. This includes information on passenger gender, age and class.

Conduct a test to see if there is an association between class and survival for *male adult* passengers.

Marking guidance

1 mark for the hypothesis; 1 marks for preliminary data exploration and visualisation, 4 marks selecting the correct data in right format, 2 marks for selection and carrying out correct test, 2 marks reporting your results.