Advanced Biometry (**MARS6300**) - Homework 6 **Name:** _____KEY_____

Distributed: Friday March 25, 2016 **Due:** Monday April 4, 2016

Instructions: Copy and paste your answers below and turn in a word file and two excel files by the end of due day via email to khyrenba@gmail.com. Please use email title "MARS 6300 hw#6" and label all files with you're a suffix including your name (e.g., MARS6300_hw6_hyrenbach). Unlabeled emails / files will be penalized 10% of points.

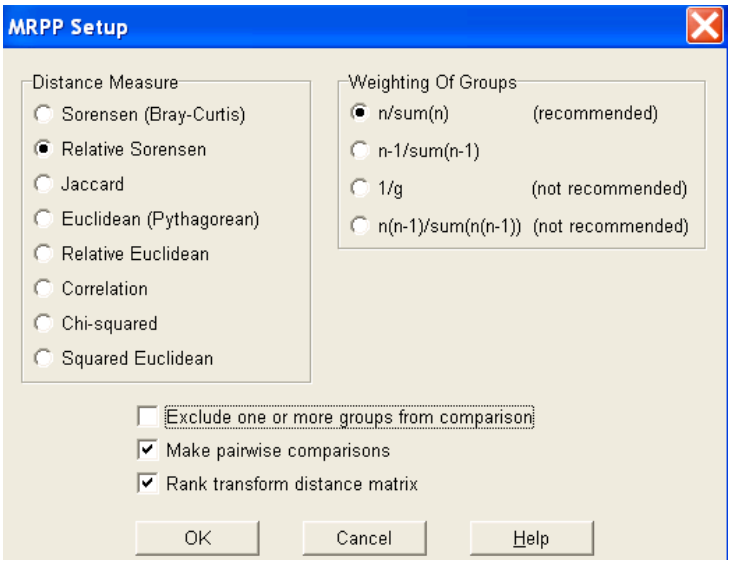
You are free to use any reference materials of your choice. While you are encouraged to work together, make sure you turn your own assignment. This homework is worth 5 points. Make sure you leave the formulas showing all of your calculations in the excel file, and explain your reasoning, to get partial credit.

The objectives of this homework are:
A) To critically evaluate the MANOVA literature
B) To perform and interpret an MRPP analysis.
C) To perform and interpret a MANOVA analysis.

To complete this homework, you will need:
- Instruction file: "BIOL6090_hw6.doc" (open with word file) – **Turn this file in**
- "Biogeography_samples.xls" and "Biogeography_groups.xls"
- One pdf: Anderson 2001

1) MRPP Analysis:

- You will perform the MRPP analysis in two ways: (i) first, you will use the grouping variable SST (in matrix 2), to identify three groups, and (ii) you will select one of the indicator species you identified in the last homework, to define groups.
- To compare the groups of samples using SST, import "Biogeography_samples.xls" as first matrix and "Biogeography_groups.xls" as matrix 2. Perform the MRPP analysis using the following set-up:



Report the following:

Test statistic: T =	-6.6533723
Observed delta =	0.28768382
Expected delta =	0.50000000

Variance of delta =	0.10183158E-02
Skewness of delta =	-0.75156943

- Interpret the Test statistic T: What does the negative sign mean?

T = (observed delta – expected delta) / (standard deviation of expected delta). The "T" value measures the effect size. Therefore, if "T" is negative, then the observed delta (average within-group distance) value must be smaller than the expected delta. The more negative the value of "T", the stronger the separation between groups.

- What is the probability (p value) ?

p = (1 + # smaller deltas) / (total # possible shufflings). It is reported that p = 0.00000805.

- Is this a significant result?

As p = 0.00000805 is much less than p = 0.05, this is a highly significant result. Therefore, this is telling us that we can reject the null hypothesis that within-group distances are the same as across-group distances. So, distances within groups are closer than in the shuffled scenarios – The groups are significantly different in composition of their communities.

- What groups are responsible for the significant pattern?

By looking at the pairwise comparisons, we get specific p-values (1 vs. 2; 1 vs. 3; and 2 vs. 3). In all 3 cases, the p-values are less than 0.05, so all three groups are different, and all 3 are responsible for this significant pattern.

Group Codes

Compared			T	A	p
1	vs.	2	-3.13788031	0.22633745	0.01041984
1	vs.	3	-5.70044950	0.34343435	0.00067128
2	vs.	3	-4.85021806	0.33475783	0.00211280

By doing just the MRPP analysis, it is impossible to tell specifically which SPECIES are responsible for any significant patterns – all the MRPP analysis can tell you is whether or not there is a significant pattern within or across groups. To determine which SPECIES are responsible for the significant pattern, a further analysis must be completed – the ISA analysis.

2) ISA Analysis:

- Perform the ISA analysis to determine which species are responsible for the significant differences you found with the MRPP test? How can you determine which species are responsible for the significant differences you found with the MRPP test?

By doing an ISA Analysis we can determine which species are found mostly in a single group and are present at most of the sites belonging to that group (indicator species). With ISA, we calculate the proportional abundance of a species in a particular group relative to the abundance of that species in all groups (abundance/density), as well as finding the proportional frequency of the species within each group (fidelity). These 2 proportions are then multiplied, then multiplied by 100 (to be expressed as a percent), and this is the indicator value for each species within each group. The highest indicator value given to a species across all groups is the indicator value of that species.

- Copy and paste the Indicator Values for the 47 species (for all three water masses) below:

Table 1. Indicator Values of 47 bird species from 16 bins using SST as the grouping variable.

INDICATOR VALUES (% of perfect indication, based on combining the above values for relative abundance and relative frequency)

					Group		
				Sequence:	1	2	3
				Identifier:	1	2	3
				Number of items:	4	5	7
	Column	Avg	Max	MaxGrp			
1	ANTE	5	14	3	0	0	14
2	AUSH	17	50	1	50	0	0
3	BBAL	24	71	3	0	0	71
4	BBSP	17	30	2	0	30	22
5	BLPT	24	71	3	0	0	71
6	BRPT	25	75	1	75	0	0
7	BRSK	19	57	3	0	0	57
8	BUPT	17	50	1	50	0	0
9	CAPT	18	51	3	0	2	51
10	COSH	7	20	2	0	20	0
11	DKSH	5	14	3	0	0	14
12	DKTE	17	50	1	50	0	0
13	DPSP	24	71	3	0	0	71
14	FFSH	16	43	2	0	43	4
15	GBSP	24	71	3	0	0	71
16	GHAL	24	71	3	0	0	71
17	GRSH	7	20	2	0	20	0
18	GWPT	30	81	2	9	81	0
19	JFPT	7	20	2	0	20	0
20	KEGU	5	14	3	0	0	14
21	KEPT	24	71	3	0	0	71
22	KESH	5	14	3	0	0	14
23	KIPN	19	57	3	0	0	57
24	LISH	14	27	1	27	8	7
25	LMSA	17	43	3	0	8	43
26	LTJA	17	48	2	1	48	2
27	MAPN	14	43	3	0	0	43
28	MSPT	25	75	1	75	0	0
29	NGPT	24	71	3	0	0	71
30	PRSP	29	86	3	0	0	86
31	ROPN	14	43	3	0	0	43
32	SAAL	7	20	2	0	20	0
33	SGPT	14	32	3	0	9	32
34	SOAL	19	57	3	0	0	57
35	SOSH	6	18	2	0	18	2
36	SPPT	17	35	3	10	5	35
37	WAAL	22	55	3	6	4	55
38	WBSP	14	43	3	0	0	43
39	WCAL	5	9	3	0	7	9
40	WCPT	30	78	3	0	13	78
41	WFSP	6	8	3	0	8	8
42	WHPT	24	71	3	0	0	71
43	WISP	27	71	3	0	10	71
44	WNPT	8	25	1	25	0	0
45	WTSH	17	50	1	50	0	0
46	WTTR	8	25	1	25	0	0
47	YNAL	23	63	3	0	5	63

```
Averages  17  47          10    8   32
```

- Using these results, classify the 47 species as having one biogeographic affinity: (Note: explain the rule you are using to make this classification below, and then list the species acronyms below, within each biogeographic classification)
  Explain your RULE here:

Groups were determined in that a species was assigned to one specific group (1, 2, or 3) if it had only a single value in one of the three groups. It was assigned to multiple groups if it had an indicator value in more than one group. You could have used other definitions that were more stringent and only considered associations with a single group.

  - Subtropical (group 1):
    AUSH, BRPT, BUPT, DKTE, MSPT, WNPT, WTSH, WTTR
  - Northern convergence (group 1 and 2):
    GWPT
  - Convergence (group 2):
    COSH, GRSH, JFPT, SAAL
  - Southern convergence (group 2 and 3):
    BBSP, CAPT, FFSH, LMSA, SGPT, SOSH, WCAL, WCPT, WFSP, WISP, YNAL
  - Sub-Antarctic (group 3):
    ANTE, BBAL, BLPT, BRSK, DKSH, DPSP, GBSP, GHAL, KEGU, KEPT, KESH, KIPN, MAPN, NGPT, PRSP, ROPN, SOAL, WBSP, WHPT
  - Cosmopolitan (widely distributed):
    LISH, LTJA, SPPT, WAAL

- Finally, paste the Montecarlo significance results below, and indicate how many of these species were significant indicators for each of the three groups you identified?

The number of species that were significant indicators for each of the 3 groups is 16 species. For subtropical, there were 3; convergence had 1 species; and sub-Antarctic had 12. This was decided by including those species with a p-value less than 0.05.

```
MONTE CARLO test of significance of observed maximum
indicator value for species
  4999 permutations.
Random number seed:        3312

---------------------------------------------------

                      Observed         IV from
```

Column	Maxgrp	Indicator Value (IV)	randomized groups Mean	S.Dev	p *	
1 ANTE	3	14.3	18.7	4.35	1.0000	
2 AUSH	1	50.0	20.4	11.56	0.0584	
3 BBAL	**3**	**71.4**	**33.3**	**14.61**	**0.0196**	
4 BBSP	2	29.7	39.1	14.39	0.7131	
5 BLPT	**3**	**71.4**	**29.6**	**13.68**	**0.0144**	
6 BRPT	**1**	**75.0**	**23.0**	**12.84**	**0.0092**	
7 BRSK	**3**	**57.1**	**27.8**	**13.49**	**0.0434**	
8 BUPT	1	50.0	23.9	9.87	0.0584	
9 CAPT	3	51.1	29.2	13.08	0.0618	
10 COSH	2	20.0	18.8	4.37	0.5653	
11 DKSH	3	14.3	18.7	4.35	1.0000	
12 DKTE	1	50.0	21.5	10.87	0.0512	
13 DPSP	**3**	**71.4**	**32.8**	**14.67**	**0.0186**	
14 FFSH	2	43.3	26.2	12.78	0.0830	
15 GBSP	**3**	**71.4**	**31.4**	**14.68**	**0.0308**	
16 GHAL	**3**	**71.4**	**29.0**	**13.26**	**0.0104**	
17 GRSH	2	20.0	18.8	4.37	0.5653	
18 GWPT	**2**	**81.2**	**33.0**	**12.72**	**0.0064**	
19 JFPT	2	20.0	18.8	4.37	0.5653	
20 KEGU	3	14.3	18.8	4.38	1.0000	
21 KEPT	**3**	**71.4**	**31.2**	**14.57**	**0.0196**	
22 KESH	3	14.3	18.8	4.38	1.0000	
23 KIPN	**3**	**57.1**	**27.1**	**13.39**	**0.0436**	
24 LISH	1	27.1	31.6	13.03	0.5929	
25 LMSA	3	42.7	31.3	12.97	0.1976	
26 LTJA	2	47.7	28.8	12.66	0.0956	
27 MAPN	3	42.9	23.4	12.68	0.1314	
28 MSPT	**1**	**75.0**	**27.0**	**12.76**	**0.0092**	
29 NGPT	**3**	**71.4**	**28.0**	**11.95**	**0.0060**	
30 PRSP	**3**	**85.7**	**30.7**	**12.54**	**0.0024**	
31 ROPN	3	42.9	28.4	12.17	0.1642	
32 SAAL	2	20.0	18.8	4.37	0.5653	
33 SGPT	3	32.3	28.9	13.32	0.3091	
34 SOAL	3	57.1	30.4	14.00	0.0590	
35 SOSH	2	17.8	22.2	10.51	0.6539	
36 SPPT	3	34.8	38.8	13.33	0.5279	
37 WAAL	3	55.2	36.0	11.02	0.0680	
38 WBSP	3	42.9	25.1	12.48	0.1646	
39 WCAL	3	9.3	20.9	11.32	1.0000	
40 WCPT	**3**	**78.2**	**38.3**	**11.16**	**0.0020**	
41 WFSP	3	8.5	20.3	11.30	1.0000	
42 WHPT	**3**	**71.4**	**30.7**	**14.30**	**0.0144**	
43 WISP	**3**	**71.1**	**40.6**	**14.15**	**0.0338**	
44 WNPT	1	25.0	18.7	4.38	0.2466	
45 WTSH	1	49.6	26.2	12.69	0.0662	
46 WTTR	1	25.0	18.7	4.38	0.2527	
47 YNAL	3	62.7	43.1	16.27	0.1562	

* proportion of randomized trials with indicator value
  equal to or exceeding the observed indicator value.
  p = (1 + number of runs >= observed)/(1 + number of randomized runs)

Maxgrp = Group identifier for group with maximum observed IV

- Question: were any of the cosmopolitan species significant indicators? Why / why not?

Cosmopolitan species = LISH, LTJA, SPPT, and WAAL. None of these were significant indicators. This is because these species are found within all three SST groups within all three biogeographic region. Therefore, the presence of these species cannot tell you which SST group or even which biogeographic region it's within. A good species indicator would likely be one that is found only within one biogeographic region, not all 3.

- To finish this exercise, redo the MRPP analysis a second time. This time, using one of the indicator species you identified in the previous homework (#5) and use the presence – absence of this species to define two groups in matrix 2. Note: label the two groups 1 (species present) and 2 (species not present). Add this information to matrix 2 ("Biogeography_groups.xls") and analyze the data in matrix 1 ("Biogeography_samples.xls". DO THIS ONLY ONCE: pick whichever species you think is the best indicator for a specific water mass.

For this MRPP, I decided to use the species "AUSH" as its indicator value for the subtropical biogeographic region is 100, and is 0 in the other two regions. I found the species "AUSH" in the biogeography samples excel sheet, and re-made the "groups" data sheet by replacing the values with 1 (species present within a bin) or 2 (species absent within a bin). As "AUSH" was only present in bins 1 and 2, these 2 bins were given the value of "1" while all other bins were given the value of "2" in the secondary matrix.

- Report the results below:

Test statistic: T =	-3.276433
Observed delta =	0.40960831
Expected delta =	0.50000000
Variance of delta =	0.76288544E-03
Skewness of delta =	-0.66266154

- Interpret the Test statistic T: What does the negative sign mean?

T = (observed delta – expected delta) / (standard deviation of expected delta). The "T" value measures the effect size. Therefore, if "T" is negative, then the observed delta (average within-group distance) value must be smaller than the expected delta. The more negative the value of "T", the stronger the separation between groups.

- What is the probability (p value) ?

p = (1 + # smaller deltas) / (total # possible shufflings). It is reported that p = 0.00429651.

- Is this a significant result?

As p = 0.00429651 is much less than p = 0.05, this is a highly significant result. Therefore, this is telling us that we can reject the null hypothesis that within-group distances are the same as across-group distances. So, distances within groups are closer than in the shuffled scenarios –

The groups (1 and 2, or AUSH presence vs. AUSH absence) are significantly different in composition of their communities.

- Are you surprised of this result? Explain why / why not?

No, I am not surprised by this result. AUSH is an indicator of the subtropical biogeographic region, and found nowhere else. Therefore, it seems likely that its presence/absence would be able to define whether or not groups are significantly different from each other. If it is present, then it must be the subtropical region. If not present, then it must be another region. The only problem is that it cannot be discerned whether it would be the convergence or sub-Antarctic region.

3) You have five groups of samples (5 in each), with the abundance of three species (see table 2 below). Enter this information into two matrices in PC-ORD (1: a data matrix; 2: a grouping matrix) and perform an MRPP test to determine which groups and species are significantly different.

TABLE 2. Test case example: abundance of three species in 25 sites divided into five clusters.

Species	Group 1					Group 2					Group 3					Group 4					Group 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3	3	3	3	3	3
B	8	8	8	8	8	4	4	4	4	4	6	6	6	6	6	4	4	2	0	0	0	0	0	0	0
C	18	18	18	18	18	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

First, calculate the indicator values of each species in each group, as the product of the "fidelity" and the "specificity" values:

	Group1	Group2	Group3	Group4	Group5
SpA - fidelity	5 / 5 = 1	5/5 =1	5/5 = 1	5/5 = 1	5/5 = 1
SpA - specificity	4 / 20 = 0.2	5/20 = 0.25	5/20 = 0.25	3/20 = 0.15	3/20 = 0.15
SpA - Indicator Value	**0.20**	**0.25**	**0.25**	**0.15**	**0.15**
SpB - fidelity	5/5 = 1	5/5 = 1	5/5 =1	3/5 = 0.6	0/5 = 0
SpB - specificity	8/20 = 0.4	4/20 = 0.2	6/20 = 0.3	2/20 = 0.1	0/20 = 0
SpB - Indicator Value	**0.40**	**0.20**	**0.30**	**0.06**	**0.00**
SpC - fidelity	5/5 = 1	5/5 = 1	0/5 = 0	0/5 = 0	0/5 = 0
SpC- specificity	18/20 = 0.9	2/20 = 0.1	0/20 = 0	0/20 = 0	0/20 = 0
SpC - Indicator Value	**0.90**	**0.10**	**0.00**	**0.00**	**0.00**

- List which samples (1, 2, 3, 4, 5) are best indicated by the three species:
  Species A: _ Groups 1, 2, 3, 4, and 5, but **groups 2 and 3** have the highest indicator values. **Therefore, species 1 (or A) is best indicated within groups 2 and 3**, but is present within all groups.

Species B: _ Groups 1, 2, 3, and 4, but **group 1** has the highest indicator value. **Therefore, species 2 (or B) is best indicated within group 1**, but is present within 4 groups.
Species C: _ Groups 1 and 2, but **group 1** has the highest indicator value. **Therefore, species 3 (or C) is best indicated within group 1**, but is also present within group 2.

- Now that you have calculated the indicator species values, perform the MRPP test, using 5 groups. Report the results below:

Test statistic: T =	-10.566117
Observed delta =	0.69230767E-01
Expected delta =	0.41333334
Variance of delta =	0.10605838E-02
Skewness of delta =	-0.78850599

- Interpret the Test statistic T: What does the negative sign mean?

T = (observed delta – expected delta) / (standard deviation of expected delta). The "T" value measures the effect size. Therefore, if "T" is negative, then the observed delta (average within-group distance) value must be smaller than the expected delta. The more negative the value of "T", the stronger the separation between groups.

- What is the probability (p value) ?

p = (1 + # smaller deltas) / (total # possible shufflings). It is reported that p = 0.00000000. It is likely that the p-value isn't actually "0" but just is rounded as it is still very small.

- Is this a significant result?

As p = 0.00000000 is much less than p = 0.05, this is a highly significant result. Therefore, this is telling us that we can reject the null hypothesis that within-group distances are the same as across-group distances. So, distances within groups are closer than in the shuffled scenarios – The groups are significantly different in composition of their communities. In fact, when comparing all groups in a pairwise comparison, all have a p-value smaller than 0.05.

Finally, perform an ISA in PC-ORD to see which species are responsible for these differences and report the following:

Species	Indicator Value	P Value
A	25.0	0.0260
B	40.0	0.0002
C	90.0	0.0002

Do your scores agree with the PC-ORD Indicator Values? Why / Why Not?

Yes, the scores we calculated above do agree with the PC-ORD indicator values. The indicator value reported by PC-ORD for each species is the same as the largest indicator value for each species within all 5 groups (Species A = 0.25 or 25; Species B = 0.40 or 40; Species C = 0.90 or 90).

4) Read and Evaluate Anderson 2001

Focus on the section "ECOLOGICAL EXAMPLES", and for each of these approaches, create a matrix showing how you would organize your samples (like it was shown in class) to obtain the correct sample design. Hint: You do not need to include the species data, just show how you would arrange each of the treatments and replicates to re-create the experimental design discussed in the particular example. Label factors and replicates in separate columns

(i) Two way factorial design:

24 Sampling Units

2 Variables

	C	C
	Factor 1 = Position	Factor 2 = Shading
SU1	1	1
SU2	1	1
SU3	1	1
SU4	1	1
SU5	1	2
SU6	1	2
SU7	1	2
SU8	1	2
SU9	1	3
SU10	1	3
SU11	1	3
SU12	1	3
SU13	2	1
SU14	2	1
SU15	2	1
SU16	2	1
SU17	2	2
SU18	2	2
SU19	2	2
SU20	2	2
SU21	2	3
SU22	2	3
SU23	2	3
SU24	2	3

Table showing a two-way factorial design for factor 1 (position – near or far from sea floor) and factor 2 (shading – shade, procedural control, or no shade). It is assumed that there are 4 replicates (4 different settlement plates) for each combination of factors 1 and 2.

(ii) Three way design, including nesting:

30 Sampling Units

3 Variables

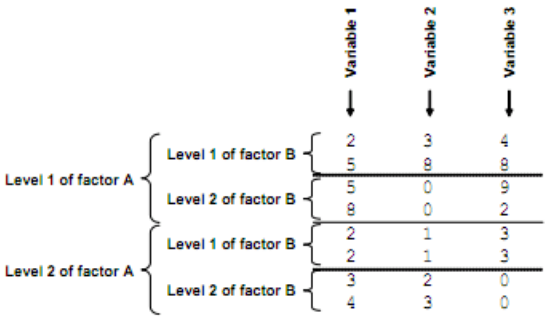
	C	C	C
	Factor 1 = Time	Factor 2 = Patch	Nest = Sticks
SU1	1	1	1
SU2	1	1	2
SU3	1	2	1
SU4	1	2	2
SU5	1	3	1
SU6	1	3	2
SU7	2	1	1
SU8	2	1	2
SU9	2	2	1
SU10	2	2	2
SU11	2	3	1
SU12	2	3	2
SU13	3	1	1
SU14	3	1	2
SU15	3	2	1
SU16	3	2	2
SU17	3	3	1
SU18	3	3	2
SU19	4	1	1
SU20	4	1	2
SU21	4	2	1
SU22	4	2	2
SU23	4	3	1
SU24	4	3	2
SU25	5	1	1
SU26	5	1	2
SU27	5	2	1
SU28	5	2	2
SU29	5	3	1
SU30	5	3	2

Table showing a three-way design, including nesting: factor 1 = time (5 periods of submersion), factor 2 = patch (3 sizes), and nesting variable = sticks (2 sticks per time x patch combination).

5) Perform MANOVA analysis:

Enter the following data into the main matrix of PC-ORD and develop a secondary matrix for the grouping of the data, using two separate groups (factors). Use Relative Sorensen and ask for group-comparisons and randomization tests.

Two-Way analysis: 2 factors A and B, each with two levels.
Note: There are 2 replicates per cell, and 3 species (variables).



Perform the test and report (cope / paste) the entire results file below.

```
******************** Permutation based MANOVA (PerMANOVA) ********************
PC-ORD, 5.10
15 Apr 2011, 13:10

Permutation-based nonparametric MANOVA calculated with method of:
Anderson, M. J.  2001.  A new method for non-parametric multivariate
     analysis of variance.  Austral Ecology 26:32-46.

HW6Q5_perMANOVA_2

        Groups were defined by values of: FactorA
                                    and: FactorB

        Main matrix has:     8 Sampling by     3 Variable

        Distance measure = Relative Sorensen


Evaluation of differences in Variable between groups.
Design: Two-way factorial
Randomization test of significance of pseudo F values
     Number of randomizations:        4999
     Random number seed:              4991 selected by time.

-----------------------------------------------------------------------
Source     d.f.        SS           MS           F           p *
-----------------------------------------------------------------------
FactorA       1    0.48406E-01  0.48406E-01   1.9269         0.225200
FactorB       1    0.19546      0.19546       7.7805         0.008600
Interac.      1    0.23075      0.23075       9.1855         0.003000
Residual      4    0.10049      0.25121E-01
Total         7    0.57510
-----------------------------------------------------------------------


Statistics from randomizations
-------------------------------------------------------------------------------------------
                                  F from randomized groups              Number
                F         ----------------------------------------     > or =
 Source        Observed         Mean         Maximum          S.Dev      observed F      p *
-------------------------------------------------------------------------------------------
FactorA         1.92690        1.36570        9.18547        0.02156       1125       0.225200
FactorB         7.78054        1.35118        9.18547        0.01619         42       0.008600
Interac.        9.18547        1.32285        9.18547        0.01180         14       0.003000
-------------------------------------------------------------------------------------------
* proportion of randomized trials with indicator value
  equal to or exceeding the observed indicator value.
  p = (1 + number of runs >= observed)/(1 + number of randomized runs)
```

```
Variance components estimated for Mixed Model II, first factor assumed fixed)
Ignore variance components if you consider the factor to have fixed effects.
COMPONENTS OF VARIANCE
----------------------------------------
Source     Variance       % of variation
FactorA                Variance component not estimated for fixed factor
FactorB    -0.88234E-02         -7.408
Interac.    0.10281              86.317
Residual    0.25121E-01          21.090
Total       0.11911             100.000
----------------------------------------

FactorA  has only two levels, so pairwise comparisons are not needed for this factor.

FactorB  has only two levels, so pairwise comparisons are not needed for this factor.
************************** PerMANOVA finished **************************
```

Report the pseudo-F values, the p values and the pair-wise tests.

-Factor A: pseudo-F = 1.9269, p = 0.2252
-Factor B: pseudo-F = 7.7805, p = 0.0086
-Interaction: pseudo-F = 9.1855, p = 0.003
-Because factor A and B have only 2 levels, pairwise comparisons are not needed for either of the factors.

Was there a significant interaction between the two factors? Why / why not? Explain.

Yes, there was a significant interaction (p = 0.003, which is less than p = 0.05). This means we must reject the null hypothesis that the interaction between the 2 factors has no effect, and it in fact does. So, the effect of Factor A on the variables depends on Factor B, and vice versa. This is because Factor B is nested within Factor A, so it makes sense that the 2 are significantly related.