UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Applied Statistics

Problem sheet 2   **Semester 2, 2020–2021**

Topic: The two-sample permutation tests

Assessed question:

1. In an experiment, the following survival times (in minutes) of the peroneal nerves of four cats and ten rabbits were recorded under a certain condition.

Cats (c_i)	25	33	43	45						
Rabbits (r_j)	15	16	16	17	20	23	28	28	35	35

   We want to do a permutation test based on sample means at $\alpha = 0.05$, to investigate whether these samples are from the same distribution, and use $\sum_{i=1}^{4} c_i$ as the test statistic.

   (a) Define your null and alternative hypotheses. [1 mark]

   (b) Why can we use $\sum_{i=1}^{4} c_i$ as the test statistic instead of $d = \bar{c} - \bar{r}$? [1 mark]

   (c) Find the p-value and state clearly what you can conclude from the result of your test. [3 marks]

   Now suppose instead we wish to carry out the Mann-Whitney U-test

   (d) Define the test hypotheses. [1 mark]

   (e) Find the observed value of the test statistic U (without using R). [2 marks]

   (f) Use R to find the (approximate) p-value and compare the result with that of part (c) for $\alpha = 0.05$. [2 marks]

Additional questions:

2. Wanting to test whether teenagers are more or less able than adults at transcribing text, an investigator randomly selects three teenagers and four adults and counts the number of transcription errors they make when presented with a standard text. The teenagers made 8, 6 and 4 errors while the adults made 7, 10, 12 and 9 errors. Which of the numbers below is the correct p-value for a relevant permutation test of equality of the means?

   (a) 1/35
   (b) 2/35
   (c) 3/35
   (d) 4/35

3. In a permutation test, the p-value is found by which of the following procedures?

   (a) Generating all possible permutations of the data assuming that the alternative hypothesis is true, and calculating the proportion of permutations that give a test statistic at least as large as that for the original data.

   (b) Generating all possible permutations of the data assuming that the null hypothesis is true, and calculating the proportion of permutations that give a test statistic that is strictly greater than that for the original data.

   (c) Generating all possible permutations of the data assuming that the null hypothesis is true, and calculating the proportion of permutations that give a test statistic at least as extreme as that for the original data.

   (d) Generating all possible permutations of the data assuming that the alternative hypothesis is true, and calculating the proportion of permutations that give a test statistic that is at least as extreme as that for the original data.

4. Reaction times were recorded for 13 individuals under the influence of either drug A or drug B, giving the following data:

Drug A	1.96	2.24	1.71	2.41	1.62	1.93	
Drug B	2.11	2.43	2.07	2.71	2.50	2.84	2.88

   (a) Carry out a two-sample permutation test, based on sample means (enumerating all permutations at least as extreme as the observed one), to assess whether on average the effects of the two drugs differ, and state your conclusions.

   (b) Apply a Mann-Whitney test Using `R` and compare the result with that of part (a).

   (c) Use `qqnorm` and `qqline` for each of the samples to see if they seem to be from normal distributions or not (although the sample sizes are relatively small). Carry out a `t.test` to compare the means of the two samples and compare the test result with the previous ones.

5. (From: Applied nonparametric statistical methods, by Sprent and Smeeton) We give below the age at death of male members of two Scottish clans buried in the Badenscallie burial ground in the Coigach district of Wester Ross, in Scotland. The data were collected in June 1987. Clan names have been changed but the records are as complete as possible for four real clans (two of them given here). There are inevitably a few missing values as names or dates were unreadable on several headstones, and indeed some headstones were missing, but these were few in number. Minor spelling variations, especially those of M, Mc, Mac, were ignored. Ages are given for complete years according to the information on the tombstones.

(a) Use `R` to conduct a Mann-Whitney test to assess the hypothesis that these may be regarded as samples from identical populations against that the hypothesis that the populations differ in location.

(b) Explain why the output includes a warning message?

(c) A normal approximation has been used for computing the p-value. What is the distribution of test statistics U_G under this approximation and under the assumption that H_0 is correct?

McBeta	0	19	22	30	31	37	55	56	66	66	67	67
	68	71	73	75	75	78	79	82	83	83	88	96
McGamma	13	13	22	26	33	33	59	72	72	72	77	78
	78	80	81	82	85	85	85	86	88			