

Applied Statistics

Problem sheet 3

Semester 2, 2020–2021

Topic: Randomisation tests

Assessed question:

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

Drug A: 2.07 1.71 2.24 1.62 2.11

Drug B: 2.41 1.96 1.93 2.71 2.50 2.84 2.88 2.43

- (a) Use R to carry out a Mann-Whitney U-test on these data. [2 Marks]
- (b) Use R to carry out a two-sample randomisation test, based on sample means, to assess whether the effects of the two drugs differ, and state your conclusions. [5 Marks]
- (c) Will the p -value calculated in part (b) differ to one calculated by enumerating all permutations at least as extreme as the observed one? Explain your answer. [3 Marks]

Additional questions:

2. In the table below, x and y represent the number of errors made in a typing test by ten randomly-selected individuals before (x) and after (y) doing a set of online typing tutorials.

Individual	1	2	3	4	5	6	7	8	9	10
x	31	20	9	28	37	26	23	12	10	34
y	23	20	14	2	10	29	11	2	8	20

We want to know whether the online tutorials affect the number of typing errors made.

- (a) Stating clearly your hypotheses, conduct a permutation test for matched pairs based on the sample means for these data with a two-sided alternative, by enumerating all permutations at least as extreme as the observed one. [3 marks]
 - (b) Now calculate the test statistic for a Wilcoxon's signed ranks test on the data. Use R to find the (approximate) p -value of the test.
 - (c) Conduct a randomisation test to do the test in part (a). Include a print out of your code and the obtained p -value in your solutions.
 - (d) Compare the p -values of the 3 tests, and if they are different, explain why?
3. The following R code describes part of a routine that can be used to perform a randomisation test of the difference in means of two samples of data. Assume that the two samples are stored in a data frame consisting of a column containing all the data (data) and a column containing an index (index) denoting the sample the data belongs to. Assume that this data frame has been attached to the R workspace.

```
> randn.fn <- function(nrand) {
+ r <- vector(length=nrand+1)
+ for (i in 1:nrand) {
+ ****
+ rmean <- mean(rdata[index=="1"])-mean(rdata[index=="2"])
+ r[i+1] <- rmean
+ }
+ r
+ }
```

Which of the following commands should replace the missing line ***** in the above code?

- (a) `rdata <- sample(data,length(data),replace=F)`
- (b) `rdata <- sample(index,length(index),replace=T)`
- (c) `rdata <- sample(data,length(data),replace=T)`
- (d) `rdata <- sample(c(-1,1),length(data),replace=T)*data`

4. The following analysis was conducted in R:

```

1: > year <- c(1,2,3,4,5,6,7,8,9,10)
2: > sales <- c(22,31,25,26,32,29,33,30,35,31)
3: > n <- length(year)
4: > teststat <- cor(year,sales)^2
5: > nr <- 9999
6: > rcor2 <- vector(length=nr+1)
7: > for (i in 1:nr) {
8: +   rsales <- sample(sales,n,replace=F)
9: +   rcor2[i+1] <- cor(year,rsales)^2
10: + }
11: > rcor2[1] <- teststat
12: > p <- length(rcor2[rcor2>=teststat])/length(rcor2)
13: > p
14: [1] 0.0261

```

- (a) Explain clearly what each line of the above code is performing (use the line numbers given on the left to identify which line you are explaining). What type of test is being conducted?
 - (b) State the null and alternative hypotheses being tested.
 - (c) What would you conclude from this analysis?
5. Three types of root-stock were used in an apple orchard grafting experiment. The data analysed below represent the extension growth (cm) four years after grafting.

```

1 > growth <- c(2569,2928,2865,3844,3027,2336,3211,3037,
+               2074,2885,2378,906,2782,3018,2383,2447,
+               2505,2315,2667,2390,3021,3085,3308,3231)
2 > type <- c(rep(1,8),rep(2,8),rep(3,8))
3 > n1 <- length(growth[type=="1"])
4 > n2 <- length(growth[type=="2"])
5 > n3 <- length(growth[type=="3"])
6 > n <- length(growth)
7 >
8 > teststat <- n1*(mean(growth[type=="1"])-mean(growth))^2 +
+               n2*(mean(growth[type=="2"])-mean(growth))^2 +
+               n3*(mean(growth[type=="3"])-mean(growth))^2
9 >
10 > multrand.function <- function(nrand) {
11 +   r <- vector(length=nrand+1)
12 +   for (i in 1:nrand){
13 +     rgrowth <- sample (growth, n, replace=F)
14 +     randstat <- n1*(mean(rgrowth[type=="1"])-mean(growth))^2 +
+                   n2*(mean(rgrowth[type=="2"])-mean(growth))^2 +
+                   n3*(mean(rgrowth[type=="3"])-mean(growth))^2
15 +     r[i+1] <- randstat}
16 +   r

```

```
+ }  
>  
12 > nrand <- 9999  
13 > results <- multrand.function(nrand)  
14 > results[1] <- teststat  
15 > results <- sort(results)  
16 > p <- length(results[results>=teststat])/length(results)  
17 > p  
[1] 0.0479
```

- (a) State what each numbered line of code does. (You do not need to write out the code — just refer to the line number.)
- (b) What test has been carried out?
- (c) What can you conclude from the result?