Exercises on Lectures 1–4

The real exercises are using the methods to analyse data, hence the practical classes. Unless told otherwise, use computer programs to help you answer the exercises for this course.

1. Suddath *et al* (1990) used magnetic resonance imaging to measure left hippocampal volumes (cm^2) of 15 pairs of monozygous ('identical') twins, one of each pair being schizophrenic. The paired measurements were

healthy	1.94	1.44	1.56	1.58	2.06	1.66	1.75	1.77	1.78	1.92	1.25	1.93	2.04	1.62	2.08
schizoid	1.27	1.63	1.47	1.39	1.93	1.26	1.71	1.67	1.28	1.85	1.02	1.34	2.02	1.59	1.97

(a) Plot the data and establish a suitable transformation.

(b) Test if schizophrenia is associated with a reduction in left hippocampal volume.

(c) If you find an association, describe it *on the original scale*.

2. Harder & Thompson (1989) observed bumblebee queens and honeybee workers visiting a species of lily: for each visit they recorded the proportion of pollen removed and the length of the visit in seconds. (There were 35 visits by bumblebees and 12 by honeybees: the dataset is `ex0328` in library `Sleuth`.)

(a) For each of the responses, test for a difference in distribution between the two species.

(b) Consider suitable transformations for the visit time from the Box–Cox family, such as log and reciprocal. Why might these be popular transformations for times?

(c) For proportions, the *logit* transformation is a good starting point. Try it. (If you don't know what it is, look it up!)

(d) Investigate the association between pollen removed and length of visit.

3. Doksum (1974) reported survival times of guinea pigs, 64 controls and 58 which received a dose of tubercle bacilli.

```
controls:
 18  36  50  52  86  87  89  91 102 108 114 114 115 118 119 120
149 160 165 166 167 167 173 178 189 209 212 216 273 278 279 292
341 355 367 380 382 421 421 432 446 455 463 474 505 545 546 569
576 590 603 607 608 621 634 637 638 641 650 663 685 688 725 735
bacilli:
 76  93  97 107 108 113 114 119 136 138 139 152 154 154 160 164
164 166 168 178 179 181 181 183 185 194 198 212 213 216 220 225
225 244 253 256 259 265 268 270 283 289 291 311 315 326 326 361
373 373 376 397 398 406 459 466 592 598
```

These data can be found as dataset `ex0211` in library `Sleuth`. Analyse these data.

Comment critically on whether what is tested by the t-test and the Mann-Whitney test is exactly what is required for this problem.

4. In the lectures you saw an example of the Mann-Whitney / Wilcoxon rank-sum test applied to times taken in a coordinate geometry test.

(a) Calculate by hand the rank-sum statistic.

(b) Calculate by computer the rank-sum statistic for a large random sample of permutations and compare to the observed value.

5. Diokno *et al* (1990) took a sample of 13,912 households from Washtenaw County, Michigan, identified 2,993 persons aged 60 or older of whom 1,956 agreed to participate in their study. Cross-tabulating the entries for married women (and omitting missing values for questions of whether they were sexually active and whether they drank coffee gave the 2×2 table

	Yes	No
Coffee drinker	15	25
Not drinker	115	70

(a) Apply the Pearson chi-squared test to this table and test the null hypothesis of no association. Are its assumptions satisfied?

(b) Would you expect the Fisher exact test to lead to similar conclusions? Apply it and find out.

6. Prove that the Box–Muller method of simulation from the normal distribution works as claimed in the notes.

7. Prove that the *rejection sampling* of simulation will always works, and that the expected number of attempts needed is M.