

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 these exercises.

1. Suddath *et al* (1990) used magnetic resonance imaging to measure left hippocampal volumes (cm²) of 15 pairs of monozygous ('identical') twins, one of whom was 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

- Plot the data and establish a suitable transformation.
- Test if schizophrenia is associated with a reduction in left hippocampal volume.
- 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.)

- For each of the responses, test for a difference in distribution between the two species.
- Consider suitable transformations for the visit time from the Box–Cox family, such as log and reciprocal. Why might reciprocal be a popular transformation for times? [Hint: how is the exponential distribution parametrized?]
- For proportions, the *logit* transformation is a good starting point. Try it. (If you don't know what it is, look it up!)

3. Doksum (1974) reported survival times of guinea pigs, 64 controls and 58 which received a dose of tubercle bacilli. The 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. One method of M-estimation is to fit by maximum likelihood a t_ν distribution. Plot the corresponding ψ function for $\nu = 5$ and $\nu = 20$.

5. 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.

6. 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) Are these data more useful for investigating whether coffee drinking affects sexual activity or whether sexual activity affects coffee drinking?
- (b) Apply the Pearson chi-squared test to this table and test the null hypothesis of no association. Are its assumptions satisfied?
- (c) Would you expect the Fisher exact test to lead to similar conclusions? Apply it and find out.

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

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