# Exercise: Transformations

Question 1

In a study of renal function in chronic obstructive airways disease (COAD), urine samples were collected from 21 COAD patients and 18 healthy controls. Albumin excretion by COAD patients and controls is shown in the figure below. The difference was significant (t=3.5, P<0.001). All analyses were done using the logarithms of the observations (Wilkinson *et al.* 1993).
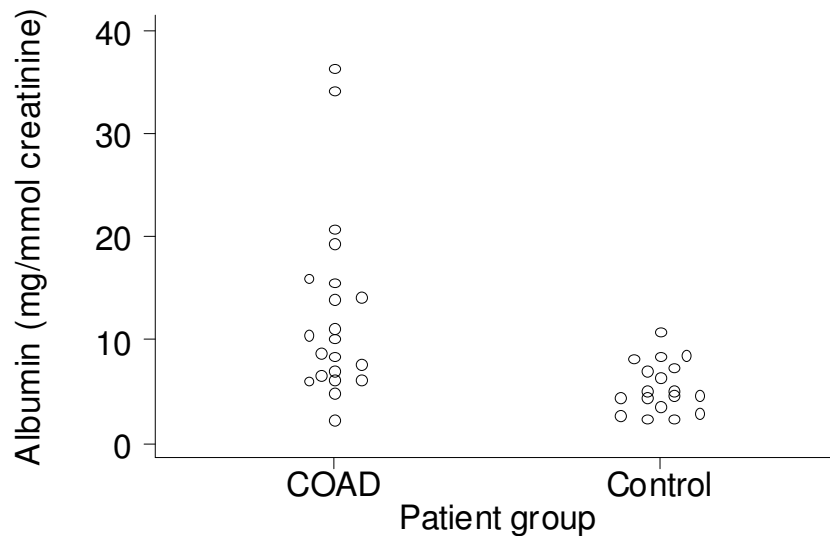

(a) Why was the logarithmic transformation of the data used?

The estimate of the difference, COAD – control, on the $\log_e$ scale was 0.720 (95% CI 0.344 to 1.096), which gives the antilog 2.05 (95% CI 1.41 to 2.99).

(b) How can we interpret the antilog of the difference and its confidence interval?

Question 2

In a study of factors related to coronary heart disease, the prevalence of arm movements was compared in 25 consecutive attenders at a chest pain clinic who had coronary heart disease (CHD) and 25 consecutive attenders at a non-cardiac clinic. Patients were fitted with a goniometer, a device which records movement. Movements were recorded during 10 minutes of the consultation. The function of the goniometer was not explained to the patient until after the reading had been taken. The data are shown in the graph below:

Distribution of no. movements
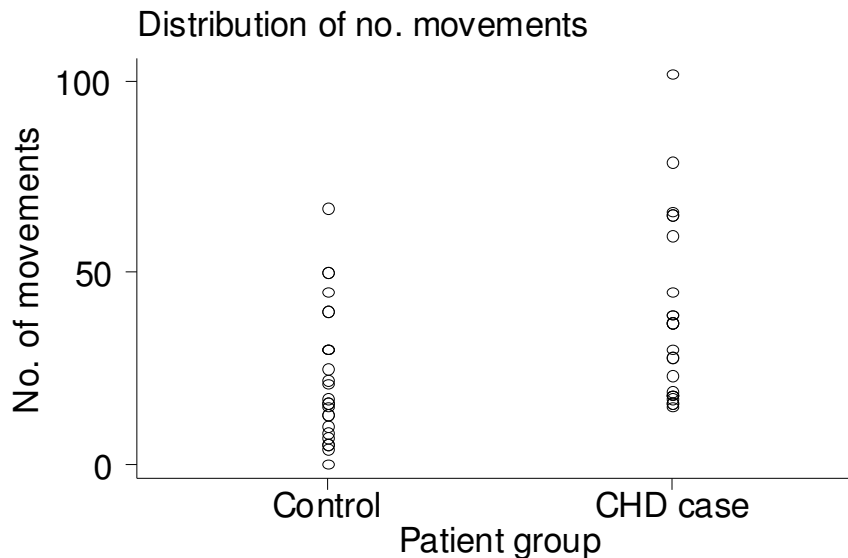
The study found that the patients with CHD made more arm movements than the controls. This difference was statistically significant using a two sample t test (P=0.01) (Rennie 1997). (In fact, Rennie described his test as one-way analysis of variance, but this is exactly the same as a two sample t test using the pooled variance when there are only two groups.)

(a) What is meant by a two sample t test and what assumptions about the data does it require?

(b) Are these assumptions likely to be met here?

(c) How could we use a transformation for these data? What problems would be caused by the subject who had zero arm movements recorded?

Question 3

The table below shows concentration of antibody to Type II Group B Streptococcus (GBS) in 20 volunteers before and after immunization (Baker *et al.* 1980).

Subject	GBS before	GBS after	Difference after—before	Subject	GBS before	GBS after	Difference after—before
1	0.4	0.4	0.0	11	0.6	12.2	11.6
2	0.4	0.5	0.1	12	0.7	1.1	0.4
3	0.4	0.5	0.1	13	0.7	1.2	0.5
4	0.4	0.9	0.5	14	0.8	0.8	0.0
5	0.5	0.5	0.0	15	0.9	1.2	0.3
6	0.5	0.5	0.0	16	0.9	1.9	1.0
7	0.5	0.5	0.0	17	1.0	0.9	−0.1
8	0.5	0.5	0.0	18	1.0	2.0	1.0
9	0.5	0.5	0.0	19	1.6	8.1	6.5
10	0.6	0.6	0.0	20	2.0	3.7	1.7

The comparison of the antibody levels was summarized in the report as of this study as '$t = 1.8$; $P>0.05$'.

(a) Comment on the presentation of the P value.

(b) Why is the paired t test not suitable for these data?

(c) Could a transformation of the antibody levels make the paired t test appropriate?

(d) What other approaches could be used?

(e) Using the table overleaf, carry out an appropriate significance test.

(f) What would this do to the conclusions of the study?

References

Rennie, A.N. (1997) Prevalence of arm movements in patients with coronary heart disease: case-control study. *British Medical Journal* **314**, 122.

Wilkinson, R., Milledge, J.S., and Landon, M.J. (1993) Microalbuminuria in chronic obstructive lung disease. *British Medical Journal* **307**, 239.

Baker, C.J., Kasper, D.L., Edwards, M.S., Schiffman, G. (1980) Influence of preimmunization antibody levels on the specificity of the immune response to related polysaccharide antigens. *New England Journal of Medicine* **303**, 173-8.

The sign test

You can use the following table to carry out the sign test. As described in Week 3, we first need the sample size, which is the number of non-zero differences. We use this because zero differences give no information about the direction of the change. We then find the test statistic. This is the number of negative differences, or the number of positive differences if this is smaller. We then find the row in the table corresponding to our sample size. The difference is significant if the test statistic less than or equal to the critical value.

Table for the sign test

Number of non-zero differences	Critical value for a two-sided test, 0.05 level
5	None
6	0
7	0
8	0
9	1
10	1
11	1
12	2
13	2
14	2
15	3
16	3
17	4
18	4
19	4
20	5