

STATISTICS PRACTICAL 2

- (1) A study was carried out to test the prevalence of side effects from the pertussis vaccine.¹ Of 339 infants who received their first injection of vaccine, 69 showed adverse reactions.
 - (a) Compute 95% and 99% confidence intervals for the probability of an adverse reaction to the vaccine.
 - (b) What do these confidence intervals mean? What assumptions go into the interpretation?
- (2) To determine the effectiveness of a new drug on the level of haemoglobin in the blood of anemic patients, 10 randomly selected patients who underwent this treatment were sampled. The table below shows the level of haemoglobin in the patients' blood before and after the treatment.

Patient	1	2	3	4	5	6	7	8	9	10
Before	11.2	9.4	9.9	9.3	8.9	8.2	10.5	8.8	10.3	9.8
After	12.9	10.8	10.3	10.9	8.5	8.9	10.4	8.5	11.2	10.1

- Carry out both the two-sample t test and the paired sample t test at a significance level of 0.05 to determine if the drug is effective in increasing the level of haemoglobin in the blood. Given the setting of the experiment, which test is more appropriate? Why?
- (3) You have a small number (2, 3, 5, or 10) of independent samples from an unknown distribution with expectation 1. You perform a T test at level 0.01 for the null hypothesis that the expectation is 1. However, the distribution that you are sampling from is not normal.
 - (a) Simulate this experiment 1000 times with each of a number of different distributions: One-fifth of a Poisson RV with parameter 5, Exponential with parameter 1, Uniform on $[0, 2]$. Under which circumstances does the test work (in the sense that the probability of rejecting a true null hypothesis is close to 0.01)? (Note: For the Poisson try computing the probability that all samples in a group will be the same. What problem would you have then? Try excluding groups where all samples are equal.)
 - (b) Compare the distribution of your simulated statistic to Student t distribution using a Q-Q plot. For the Poisson case you will have to exclude (Note: The command `qqplot` makes a Q-Q plot for comparing two different data sets. One of them can be a large sample from the correct T distribution.)
 - (c) Try to formulate a guess about which properties of a distribution are required for the t test to be accurate, or for it to be conservative.
 - (4) The table below shows measurements of height and forced expiratory volume in one second (FEV) in a sample of male medical students.

¹Published as "Whooping-cough vaccination: An assessment," Miller et al., *The Lancet*, 1974. Described in Samuels and Witmer, *Statistics for Life Sciences*, p.212.

Height (cm)	FEV (litres)
164.0	3.5
170.4	3.2
171.3	3.2
172.0	3.8
176.0	3.8
177.0	5.4
178.0	3.0
181.0	4.0
183.7	4.7

- (a) Represent the data graphically.
 - (b) Calculate the sample correlation.
 - (c) Perform a linear regression and include the fitted line in your graph.
 - (d) Test whether the slope of the regression line differs significantly from zero.
- (5) The following table gives the observed counts in 1-second intervals of alpha particles emitted from a radioactive source. Use a χ^2 test at the 5% level to assess the goodness of fit to the Poisson distribution.

n	Observed
0	5267
1	4436
2	1800
3	534
4	111
5+	21