Hypothesis Testing in Action

Mark Muldoon

Departments of Mathematics and

Optometry & Neuroscience

UMIST

http://www.ma.umist.ac.uk/mrm/Teaching/2P1/

Overview

Today we'll examine three data sets and use hypothesis tests to explore them.

- **Differences in proportions:** The Boston aspirin study
- **Differences in means:** the t-Tests and H.H. Koh's macular pigment data

The Boston aspirin study

In a famous and very large study during the 1980's, several hospitals in the Boston area worked together to conduct a placebo controlled, double-blind study of the efficacy of aspirin in preventing heart attacks. The results were:

Group	N-attacks	N Patients
Aspirin	104	11037
Placebo	189	11034

Is this an important difference?

Setup

The first question to ask is: How likely is this difference to have arisen by chance? We begin with a hypothesis test—based on a z-score—that addresses this question.

Null Hypothesis The two proportions are the same.

Alternative Hypothesis Either "The two proportions *differ*" (two-sided test) or "The placebo group had *more* attacks" (one-sided test).

Differences of proportions, large samples

The ingredients for this test are two experimentally observed proportions: $p_1 = (r_1/n_1)$ and $p_2 = (r_2/n_2)$.

(a) As the null hypothesis is that the proportions are the same, combine the data to get a single estimate of the underlying proportion:

$$p = \frac{r_1 + r_2}{n_1 + n_2}$$

(b) Estimate the standard error of the difference between the two measured proportions:

$$\text{SE} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Differences of proportions, continued

(c) Compute

$$\begin{aligned} z &= \frac{p_1 - p_2}{\text{SE}} \\ &= \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \end{aligned}$$

(d) Consult the table for the standard normal.

Application to the aspirin data

(a) Under the null hypothesis our best estimate for p is

$$p \;=\; (189+104)/(11034+11037) \;\approx\; 0.0133$$

(b) The standard error of the difference is then

$$\text{SE} \;=\; \sqrt{p(1-p)\left(\frac{1}{11034}+\frac{1}{11037}\right)} \;\approx\; 0.00154$$

(c) The z-score is

$$z \;=\; \frac{(189/11034)-(104/11037)}{0.00154} \;\approx\; 5.00$$

Application to the aspirin data

(d) This is a massively implausible z-score: we can reject the null hypothesis in favour of the alternative that the aspirin group has fewer heart attacks with confidence $\geq$ 99.99%.

Visual pigments and macular disease

The next two examples involve a novel instrument developed in the Dept. of Optometry here at UMIST. It measures the concentration of certain visual pigments and I'll refer to the data as Macular Pigment Optical Density (MPOD).

The data come from two groups: patients suffering from macular disease and healthy control subjects.

- Raw data are 10 total measurements per subject, collected in two sessions of 5 measurements and with around a 30 minute break in between.
- The MPOD is the difference between measurements taken at central fi xation and another in the periphery (5 degree visual angle).
- All measurements are on healthy eyes even among the patients, each of whom who had only one diseased eye.

Are the two groups different?

Patients		Controls	
MPOD	SEM	MPOD	SEM
0.189	0.060	0.377	0.119
0.301	0.095	0.26	0.082
0.072	0.023	0.161	0.051
0.242	0.077	0.119	0.038
0.271	0.086	0.295	0.093
0.32	0.101	0.055	0.017
0.409	0.129	0.037	0.012
0.279	0.088	0.179	0.057
0.556	0.176	0.453	0.143
$m_x \approx 0.293$		$m_y \approx 0.215$	
$s_x \approx 0.135$		$s_y \approx 0.142$	

Testing for differences

If anything, the patients seem to have more pigment than the controls. Is this apparent difference significant? Test with a new hypothesis test, the *Two Sample t-test*, designed for differences in the means of small samples.

Null Hypothesis MPOD for Patients and Controls are drawn from the same (normal) distribution.

Alternative Hypothesis MPOD for the two groups drawn from normal distributions with same variance, but different mean.

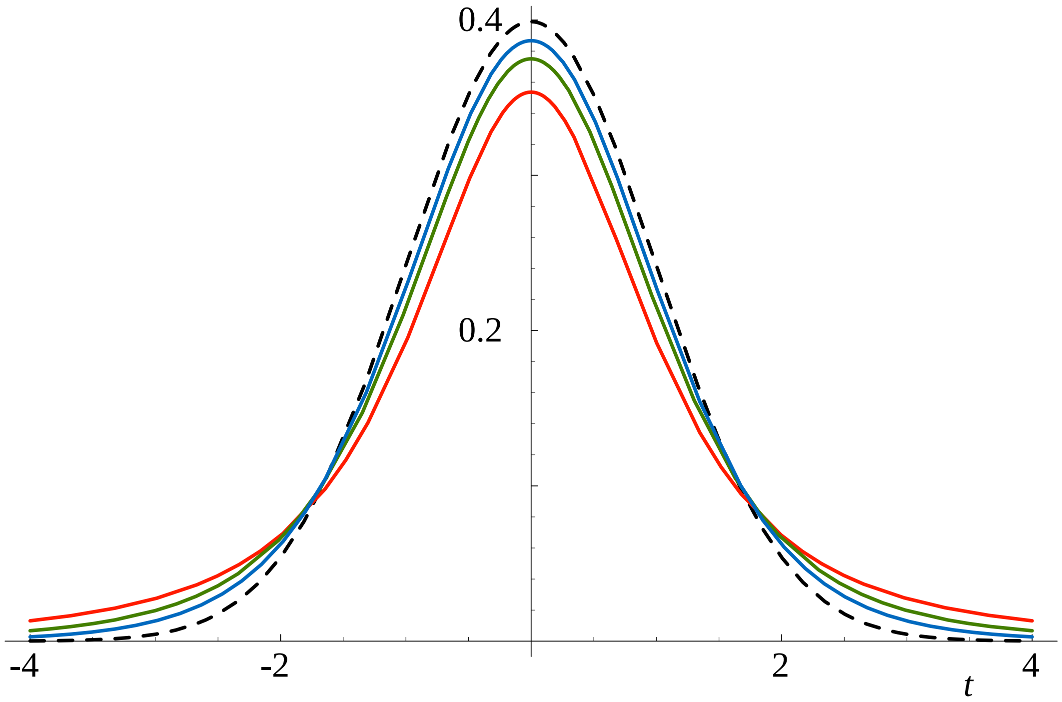
This will involve a two-sided test based on a new statistic, t.

Folklore

The t-test was developed by W.S. Gosset (1876-1937), a statistician who worked for the Guiness brewing company. Employees of the firm were not allowed to publish under their own names so he wrote under the pseudonym 'Student'. The t-statistic is:

- similar to a z-score, but is applicable when the sample is too small to assume that s_x^2 and s_y^2 provide good estimates of the variances;
- this advantage comes at a small cost: the t-distribution (and hence the tables one consults to use it) are more complicated than those for the normal distribution;
- depends on the size of the samples—when this grows large the distribution of t tends to the normal.

Distribution of t



Student's t-distribution for $\nu = 2$, 4 and 8. The dashed curve at the top is the standard normal distribution (μ = 0, σ = 1).

Computing t for two samples

The ingredients are two of lists of numbers, say, $\{x_1, x_2, \ldots, x_{N_x}\}$ and $\{y_1, y_2, \ldots, y_{N_y}\}$.

(a) Computes the two sample means, m_x and m_y. Recall that, for example,

$$m_x = \frac{\sum_{j=1}^{N_x} x_j}{N_x}.$$

(b) Computes the two standard deviations, s_x and s_y. Recall that, for example,

$$s_y^2 = \frac{\sum_{j=1}^{N_y} (y_j - m_y)^2}{(N_y - 1)}.$$

Two-sample t continued

(c) Computes the *pooled standard deviation*, s, which satisfies

$$s^2 = \frac{(N_x - 1)s_x^2 + (N_y - 1)s_y^2}{(N_x - 1) + (N_y - 1)}.$$

(d) Last, one computes

$$t = \frac{m_x - m_y}{s} \sqrt{\frac{N_x N_y}{N_x + N_y}}.$$

(e) Consult the t-table for $\nu = N_x + N_y - 2$ degrees of freedom.

Testing the MPOD data

Working through the recipe, $N_x = N_y = 9$ and:

(a) Patients had $m_x = 0.293$; Controls had $m_y = 0.215$.

(b) Patients had $s_x = 0.135$; Controls had $s_y = 0.142$.

(c) The pooled variance is thus

$$\begin{aligned} s^2 &= \frac{(N_x - 1)s_x^2 + (N_y - 1)s_y^2}{(N_x - 1) + (N_y - 1)} \\ &= \frac{8(0.135)^2 + 8(0.142)^2}{16} \\ &\approx 0.0192 \end{aligned}$$

Testing the MPOD data, continued

(d) Thus

$$
\begin{aligned}
t &= \frac{m_x - m_y}{s}\sqrt{\frac{N_x N_y}{N_x + N_y}} \\
&= \frac{0.293 - 0.215}{0.138}\sqrt{\frac{81}{18}} \\
&\approx 1.19
\end{aligned}
$$

(e) This is smaller than the critical value, 2.120, for a two-sided test with $\nu = 16$ degrees of freedom at 95% confidence.

We cannot reject the null hypothesis.

Paired sample design

- The considerable variation *within* the groups may make it hard to see whether there is much systematic difference *between* groups.
- Design a new type of study in which Patients and Controls are matched for age, gender, eye (left or right), iris colour and smoking habits.
- Compare with the *Paired Sample t-test*.

Computing t for paired samples

The only ingredient is a list of N pairs of numbers $\{(x_1, y_1), \ldots, (x_N, y_N)\}$.

- Null hypothesis is that the two members of each pair are drawn from normal distributions having the same mean.
- All the distributions for the x's are assumed to share the same variance as are all the y's, but the variance shared by the x's need not equal that shared by the y's.

Computing t for paired samples, continued

(a) Compute the differences $\delta_j = (x_j - y_j)$;

(b) Compute the mean of the differences

$$m = \frac{\sum_{j=1}^{N} \delta_j}{N};$$

(c) Estimate the variance of the differences

$$s^2 = \frac{\sum_{j=1}^{N} (\delta_j - m)^2}{N-1};$$

Computing t for paired samples, continued

(d) Computes the paired-sample t-statistic

$$t = \frac{m\sqrt{N}}{s}.$$

(e) Check against critical values in the t-table, here using $\nu = N - 1$ degrees of freedom.

Paired MPOD data, differences

MPOD		
Control	Patient	δ
0.191	-0.035	0.226
0.167	-0.014	0.181
0.029	-0.027	0.056
0.282	0.191	0.091
0.320	0.266	0.054
0.444	0.346	0.098
0.724	0.027	0.698
0.369	0.176	0.193
0.382	0.320	0.063

Paired MPOD data, differences

The mean difference is $m = 0.184$ with standard deviation $s = 0.203$. This leads to

$$\begin{aligned} t &= \frac{m\sqrt{N}}{s} \\ &= \frac{0.184\sqrt{9}}{0.203} \\ &\approx 2.723 \end{aligned}$$

This far exceeds the critical value, 2.306, for a two-sided test at the 95% confidence level ($\alpha = 0.025$, $\nu = 8$). We can reject the null hypothesis and conclude that the difference is nonzero.