Hypothesis Tests

Often investigators wish to ask questions about populations on the basis of their samples

E.g. Are the means equal?

Asking if $m_1 = m_2$ is trivial but unhelpful

Asking if $\mu_1 = \mu_2$ is helpful but more challenging

An Example

Trial of intrathecal Fentanly vs intrathecal saline for relief of post-thoracotomy pain

Ten patients randomised to each treatment

Outcome is PEFR (l/min) one hour after admission to HDU

Population mean PEFR in Fentanyl group is μ_F and saline group is μ_S and interest focuses on $\tau = \mu_F - \mu_S$

(full account in *British Journal of Anaesthesia* 1995, 75, 19-22)

Hypothesis test I: purpose

What do we know that is relevant?

$m_F - m_S$ contains information on $\mu_F - \mu_S$

The SE of $m_F - m_S$ gives an indication of how far the estimate strays from the population value.

Could use confidence interval – see appendix for links with hypothesis test

Hypothesis test is designed to address the question of whether an observed difference is ‘due to chance’

Hypothesis test II: null hypothesis

Approach is

1. assume that the observed difference is due to chance encapsulated in the *null hypothesis*

2. assess how likely this is in the light of the data

Note: even if procedure concludes difference *could* be due to chance, this does not mean that *it is* due to chance.

Likelihood in 2 is the *P*-value

Hypothesis tests III: interpreting the result

Result is that either

i) an event with probability P has occurred

OR

ii) the null hypothesis is false

Small P (say < 0.05) moves interpretation towards ii) whereas larger values of P suggest i).

Application to Pain Relief Trial

(Note: it is assumed that PEFR has a Normal distribution)

Assume $\mu_F - \mu_S = 0$ *the null hypothesis*

Under this assumption, what values of $m_F - m_S$ are plausible?

We know from properties of distribution of sample means that about 95% of means are between –2SE, 2SE, or

$$\frac{m_F - m_S}{SE} = t \text{ statistic}$$

has a 95% chance of being between –2 and 2.

Calculations

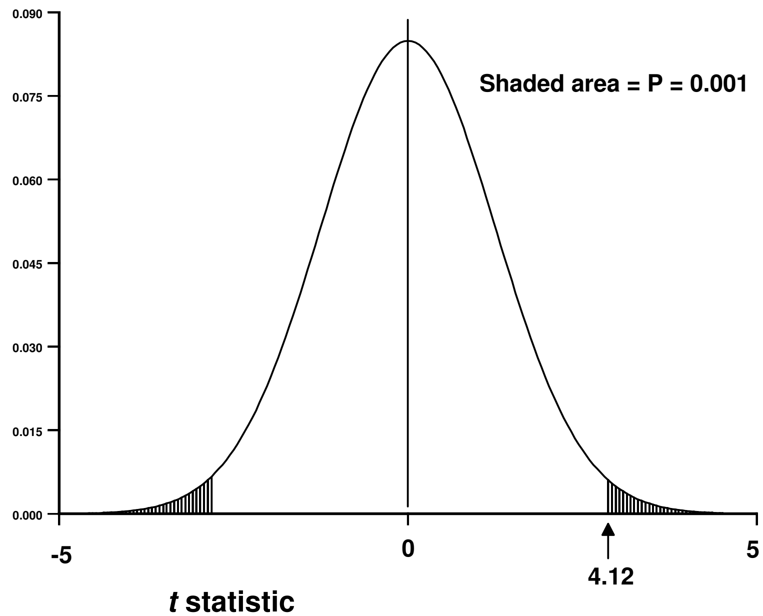
Group	Size	Mean PEFR (l/min)	SD (l/min)
Fentanyl	10	235	47
Saline	10	137	58

Will see that SE = 23.76 l/min

Difference in means = 98 l/min, so t = 98/23.76 = 4.12

Outside –2 to 2 so above suggests that an unusual event (i.e. small P-value) has occurred. Can we be more precise?

The *t* distribution



If null hypothesis is true then the range of *t* values will follow the above distribution

Proportion of distribution more extreme than 4.12 is shaded

Result of t-test is $P = 0.001$

Strong evidence against the null hypothesis

What if we had obtained, e.g., $P = 0.45$?

This would say that the sort of difference we observed would be quite commonplace if the null hypothesis were true

This is *not* the same as saying the null hypothesis is true

Data are compatible with $\mu_F - \mu_S = 0$ but also many other values.

Which ones? $\rightarrow$ use confidence interval, see appendix I

Types of *t* test

Two main types of *t*-test with a variety of names

Paired	Unpaired
One sample	Two sample
Dependent	Independent

Both forms based on $\frac{m_F - m_S}{SE}$

SE calculated differently, so as to reflect difference in structure of data

The Unpaired Test

Compares means on basis of two unrelated samples

Fentanyl vs saline comparison is of this form

Assumptions:Both populations Normal
Equal Standard deviations

Second assumption restricting?

a) often true
b) when not, often due to skewness, see later
c) Normal distribution characterised by mean and SD, so assessing means alone not always appropriate

SE is $\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}$ where σ is the common standard deviation

The Paired Test

Consider another aspect of pain study:

Standard therapy (PCA morphine) at admission to HDU and 1 hour later

PEFR on admission to HDU	PEFR one hour post admission
100	110
80	60
180	160
60	80
210	200
130	80
80	90
80	60
120	80
250	280

Here each observation in one sample is linked (paired) to an observation in other sample

Should not use an unpaired test

PEFR on admission to HDU (I)	PEFR one hour post admission (II)	(II) randomly re-ordered	Change
100	110	90	10
80	60	80	-20
180	160	160	-20
60	80	110	20
210	200	280	-10
130	80	60	-50
80	90	200	10
80	60	60	-20
120	80	80	-40
250	280	80	30

Comparison of column 1 vs column 2 is same as column1 vs column 3

This is not appropriate

Paired test analyses *differences* ('change' column)

Assumes that difference have a Normal distribution

t statistic is $\bar{d} / SE(d) = \bar{d} / \left(s_d / \sqrt{n}\right)$.

Example of Paired Test

	PEFR on admission to HDU (I)	PEFR one hour post admission (II)	Change
	100	110	10
	80	60	-20
	180	160	-20
	60	80	20
	210	200	-10
	130	80	-50
	80	90	10
	80	60	-20
	120	80	-40
	250	280	30
Mean	129	120	-9
SD	64	72	26

Standard deviation of change much smaller than of individual observation

Taking difference has removed ‘inter-patient’ variation

SE of difference is $26/\sqrt{10} = 8.2$, $t = -9/8.2 = -1.09$, $P = 0.30$