Testing means

Krzysztof Podgórski
Department of Mathematics and Statistics
University of Limerick

September 24, 2009

Linear combination

We consider the so called null hypothesis:

$$H_0 : \quad \mu = \mu_0,$$

where μ is unknown but true value of the mean while μ_0 is a known hypothesized value of the true mean.

Linear combination

We consider the so called null hypothesis:

$$H_0 : \quad \mu = \mu_0,$$

where μ is unknown but true value of the mean while μ_0 is a known hypothesized value of the true mean.
The **significance level** is a small value close to zero often expressed in percentage for example 5%. It represent the percentage of mistakenly rejecting H_0 hypothesis that we can afford.

Test and example

In order to decide whether the difference between $\bar{x}$ and μ is significant, that is to test H_0: population mean $= \mu$, the statistic t is calculated:

$$t = (\bar{x} - \mu)\sqrt{n}/s \qquad (3.1)$$

where $\bar{x}$ = sample mean, s = sample standard deviation and n = sample size.

If $|t|$ (i.e. the calculated value of t without regard to sign) exceeds a certain **critical value** then the null hypothesis is rejected. The critical value of t for a particular significance level can be found from Table A.2. For example, for a sample size of 10 (i.e. 9 degrees of freedom) and a significance level of 0.01, the critical value is $t_9 = 3.25$, where, as in Chapter 2, the subscript is used to denote the number of degrees of freedom.

Example 3.2.1

In a new method for determining selenourea in water, the following values were obtained for tap water samples spiked with 50 ng ml^{-1} of selenourea:

50.4, 50.7, 49.1, 49.0, 51.1 ng ml^{-1}

(Aller, A. J. and Robles, L. C. 1998. *Analyst* 123: 919).

Is there any evidence of systematic error?

The mean of these values is 50.06 and the standard deviation is 0.956. Adopting the null hypothesis that there is no systematic error, i.e. $\mu = 50$, and using equation (3.1) gives

$$t = \frac{(50.06 - 50)\sqrt{5}}{0.956} = 0.14$$

From Table A.2, the critical value is $t_4 = 2.78$ ($P = 0.05$). Since the observed value of $|t|$ is less than the critical value the null hypothesis is retained: there is no evidence of systematic error. Note again that this does not mean that there are no systematic errors, only that they have not been demonstrated.

```
x=c(50.4,50.7,49.1,49.0,51.1)
mean(x)
sd(x)

t.test(x,mu=50)
# One Sample t-test

#data:  x
#t = 0.1404, df = 4, p-value = 0.89
#alternative hypothesis: true mean
#95 percent confidence interval:
# 48.87358 51.24642
#sample estimates:
#mean of x
#    50.06
```

Comparison of two experimental means

We consider a null hypothesis of the form:

$$H_0 : \quad \mu_1 = \mu_2,$$

where μ_1 and μ_2 are hypothesized values of the means for two populations.

Comparison of two experimental means

We consider a null hypothesis of the form:

$$H_0 : \quad \mu_1 = \mu_2,$$

where μ_1 and μ_2 are hypothesized values of the means for two populations.
The goal is to determine if there is enough support in the data to claim that the means are not equal.

The procedure

Another way in which the results of a new analytical method may be tested is by comparing them with those obtained by using a second (perhaps a reference) method. In this case we have two sample means $\bar{x}_1$ and $\bar{x}_2$. Taking the null hypothesis that the two methods give the same result, that is H_0: $\mu_1 = \mu_2$, we need to test whether $(\bar{x}_1 - \bar{x}_2)$ differs significantly from zero. If the two samples have standard deviations which are not significantly different (see Section 3.5 for a method of testing this assumption), a pooled estimate, s, of the standard deviation can be calculated from the two individual standard deviations s_1 and s_2.

In order to decide whether the difference between two sample means $\bar{x}_1$ and $\bar{x}_2$ is significant, that is to test the null hypothesis, H_0: $\mu_1 = \mu_2$, the statistic t is calculated:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3.2}$$

where s is calculated from:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \tag{3.3}$$

and t has $n_1 + n_2 - 2$ degrees of freedom.

This method assumes that the samples are drawn from populations with equal standard deviations.

Example One

In a comparison of two methods for the determination of chromium in rye grass, the following results (mg kg^{-1} Cr) were obtained:

Method 1: mean = 1.48; standard deviation 0.28
Method 2: mean = 2.33; standard deviation 0.31

For each method five determinations were made.
(Sahuquillo, A., Rubio, R. and Rauret, G. 1999. *Analyst* 124: 1)

Do these two methods give results having means which differ significantly?

The null hypothesis adopted is that the means of the results given by the two methods are equal. From equation (3.3), the pooled value of the standard deviation is given by:

$$s^2 = ([4 \times 0.28^2]) + [4 \times 0.31]^2)/(5 + 5 - 2) = 0.0873$$

$$s = 0.295$$

From equation (3.2):

$$t = \frac{2.33 - 1.48}{0.295\sqrt{\frac{1}{5} + \frac{1}{5}}} = 4.56$$

There are 8 degrees of freedom, so (Table A.2) the critical value $t_8 = 2.31$ ($P = 0.05$): since the experimental value of $|t|$ is greater than this the difference between the two results is significant at the 5% level and the null hypothesis is rejected. In fact since the critical value of t for $P = 0.01$ is about 3.36, the difference is significant at the 1% level. In other words, if the null hypothesis is true the probability of such a large difference arising by chance is less than 1 in 100.

P-value - the observed significance

We base our inference on

$$\|t\| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the quantile of t-distribution. The distribution of t is given by the graph

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
```

P-value - the observed significance

We base our inference on

$$\|t\| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the quantile of t-distribution. The distribution of t is given by the graph

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
```

- So we reject H_0 hypothesis if the observed t falls to the tails of distribution.

P-value - the observed significance

We base our inference on

$$\|t\| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the quantile of t-distribution. The distribution of t is given by the graph

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
```

- So we reject H_0 hypothesis if the observed t falls to the tails of distribution.
- One can ask how much is left in the tails of distribution if the observed value is used instead of the quantile.

P-value - the observed significance

We base our inference on

$$\|t\| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the quantile of t-distribution. The distribution of t is given by the graph

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
```

- So we reject H_0 hypothesis if the observed t falls to the tails of distribution.
- One can ask how much is left in the tails of distribution if the observed value is used instead of the quantile.
- This value is called the observed significance or p-value.

P-value - the observed significance

We base our inference on

$$\|t\| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the quantile of t-distribution. The distribution of t is given by the graph

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
```

- So we reject H_0 hypothesis if the observed t falls to the tails of distribution.
- One can ask how much is left in the tails of distribution if the observed value is used instead of the quantile.
- This value is called the observed significance or p-value.
- We reject H_0 if p-value is smaller than α.

Example One in R

```
t=seq(-4,4,by=0.01)
plot(t,dt(t,8))
2*pt(4.56,8,lower.tail = FALSE)
```

Example Two and analysis in **R**

In a series of experiments on the determination of tin in foodstuffs, samples were boiled with hydrochloric acid under reflux for different times. Some of the results are shown below:

Refluxing time (min)	Tin found (mg kg^{-1})
30	55, 57, 59, 56, 56, 59
75	57, 55, 58, 59, 59, 59

(Analytical Methods Committee. 1983. *Analyst* 108: 109)

Does the mean amount of tin found differ significantly for the two boiling times?

The mean and variance (square of the standard deviation) for the two times are:

$$30 \text{ min} \quad \bar{x}_1 = 57.00 \quad s_1^2 = 2.80$$
$$75 \text{ min} \quad \bar{x}_2 = 57.83 \quad s_2^2 = 2.57$$

The null hypothesis is adopted that boiling has no effect on the amount of tin found. By equation (3.3), the pooled value for the variance is given by:

$$s^2 = (5 \times 2.80 + 5 \times 2.57)/10 = 2.685$$
$$s = 1.64$$

From equation (3.2):

$$t = \frac{57.00 - 57.83}{1.64\sqrt{\frac{1}{6} + \frac{1}{6}}}$$
$$= -0.88$$

There are 10 degrees of freedom so the critical value is $t_{10} = 2.23$ ($P = 0.05$). The observed value of $|t|$ ($= 0.88$) is less than the critical value so the null hypothesis is retained: there is no evidence that the length of boiling time affects the recovery rate.

The table below shows the result of performing this calculation

```
xone=c(55,57,59,56,56,59)
xtwo=c(57,55,58,59,59,59)

t.test(xone,xtwo,var.equal=
```

Non-equal variances case - approximate formula

In order to test H_0: $\mu_1 = \mu_2$ when it cannot be assumed that the two samples come from populations with equal standard deviations, the statistic t is calculated, where

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{3.4}$$

with

$$\text{degrees of freedom} = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_1}\right)^2}{\left(\dfrac{s_1^4}{n_1^2(n_1 - 1)} + \dfrac{s_2^4}{n_2^2(n_2 - 1)}\right)} \tag{3.5}$$

with the value obtained being truncated to an integer.

Non-equal variances case - Example

The data below give the concentration of thiol (mM) in the blood lysate of the blood of two groups of volunteers, the first group being 'normal' and the second suffering from rheumatoid arthritis:

Normal:	1.84,	1.92,	1.94,	1.92,	1.85,	1.91,	2.07
Rheumatoid:	2.81,	4.06,	3.62,	3.27,	3.27,	3.76	

(Banford, J. C., Brown, D. H., McConnell, A. A., McNeil, C. J., Smith, W. E., Hazelton, R. A. and Sturrock, R. D. 1983. *Analyst* 107: 195)

Non-equal variances case - Example cont.

The null hypothesis adopted is that the mean concentration of thiol is the same for the two groups.

The reader can check that:

$$n_1 = 7 \quad \bar{x}_1 = 1.921 \quad s_1 = 0.076$$

$$n_2 = 6 \quad \bar{x}_2 = 3.465 \quad s_2 = 0.440$$

Substitution in equation (3.4) gives $t = -8.48$ and substitution in equation (3.5) gives 5.3, which is truncated to 5. The critical value is $t_5 = 4.03$ ($P = 0.01$) so the null hypothesis is rejected: there is sufficient evidence to say that the mean concentration of thiol differs between the groups.

```
xone=c(1.84,1.92,1.94,1.92,1.85,1.91,2.07)
xtwo=c(2.81,4.06,3.62,3.27,3.27,3.76)
t.test(xone,xtwo)
```

Paired t-test

Table 3.1 Example of paired data

Batch	UV spectrometric assay	Near-infrared reflectance spectroscopy
1	84.63	83.15
2	84.38	83.72
3	84.08	83.84
4	84.41	84.20
5	83.82	83.92
6	83.55	84.16
7	83.92	84.02
8	83.69	83.60
9	84.06	84.13
10	84.03	84.24

(Trafford, A. D., Jee, R. D., Moffat, A. C. and Graham, P. 1999. *Analyst* 124: 163)

mean $\mu_d = 0$. In order to test the null hypothesis, we test whether $\bar{d}$ differs significantly from 0 using the statistic t.

To test whether n paired results are drawn from the same population, that