

Chemometrics

MA4605

Week 4. Lecture 7. Paired t-tests

September 26, 2011

Example 3.4.1

It frequently happens that 2 methods of analysis are compared by studying test samples containing different amount of analyte. The table below contains paracetamol concentration in tablets by two different methods. Ten tablets from 10 different batches are measured by both methods.

Normal	UV	Near-infrared
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

Paired t-test in R

There is variation between measurements due to random errors, but also due to the different methods. We wish to know if the 2 methods produce significantly different results.

```
x <- c(84.63, 84.38, 84.08, 84.41, 83.82, 83.55, 83.92, 83.69, 84.06, 84.03)
```

```
y <- c(83.15, 83.72, 83.84, 84.2, 83.92, 84.16, 84.02, 83.6, 84.13, 84.24)
```

The `> t.test(x, y)` for comparing two means from independent samples does not apply. We must specify that the two samples are paired:

```
> t.test(x, y, paired = TRUE)
```

We test the hypothesis that the two methods are similar.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

Paired t-test

data: x and y

$t = 0.8821$, $df = 9$, $p\text{-value} = 0.4007$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.2487527 0.5667527

sample estimates:

mean of the differences

0.159

The p-value produced by this command is 0.4007, hence we accept the null hypothesis $H_0 : \mu_1 = \mu_2$

Testing for variances

The significance tests used for comparing means request information on whether the variances are equal or not. Before we compare the means μ_1 and μ_2 of two groups, we need to compare the two variances σ_1^2 and σ_2^2 .

The null hypothesis is that the two samples are extracted from populations with similar variances, hence

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

the statistics F is calculated

$$F = \frac{s_1^2}{s_2^2}$$

where s_1 is the larger of the two standard deviations and s_2 is the smaller one, such that their ratio is always > 1 .

- If F is close to 1, then the two variances have similar values and the null hypothesis is accepted.
- If F is much greater than 1, in fact just greater than the critical value $F_{n_1-1; n_2-1}$, the null hypothesis is rejected.
- We used to read the critical value $F_{n_1-1; n_2-1}$ from the table with the F-distribution (Fisher distribution) with (n_1-1) and (n_2-1) degrees of freedom.
- Now we use the `var.test(x, y)` command, where x and y are two vectors that contain the two samples we are comparing.

Problem 4. Chapter 3 The following data give the recovery of bromide from spiked samples of vegetable matter.

Tomato	777,790,759,790, 770,758,765
Cucumber	782,773,778,765,789,797,782

Test whether the recoveries from the two vegetables have variances which differ significantly.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

Test whether the mean recovery rates differ significantly.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

```
> tomato <- c(777, 790, 759, 790, 770, 758, 765)
> cucumber <- c(782, 773, 778, 765, 789, 797, 782)
> var.test(tomato, cucumber)
```

F test to compare two variances

data: tomato and cucumber

F = 1.6708, num df = 6, denom df = 6, p-value = 0.5485

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2870848 9.7234388

sample estimates:

ratio of variances 1.670764

The p-value produced is 0.5485, hence we accept the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$.

Test whether the mean recovery rates differ significantly.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

> t.test(tomato, cucumber, var.equal = TRUE)

Two Sample t-test

data: tomato and cucumber

t = -1.2657, df = 12, p-value = 0.2296

alternative hypothesis: true difference in means is not equal to
0

95 percent confidence interval:

-22.159883 5.874168

sample estimates:

mean of x mean of y

772.7143 780.8571

The p-value produced is 0.2311, hence we accept the null
hypothesis $H_0 : \mu_1 = \mu_2$.