Chemometrics

MA4605

Week 3. Lecture 6. Significance tests

September 20, 2011

Significance tests

We decide if the difference between a measured value and an expected value can be accounted by random error,using a statistical test called the **significance test.**
Significance tests are widely used in the evaluation of the experimental results.

Hypothesis

- Making a significance test we are testing the truth of a hypothesis.

Hypothesis

- Making a significance test we are testing the truth of a hypothesis.
- A **hypothesis** is a theory or statement whose truth has yet to be proven or disproved.

Hypothesis

- Making a significance test we are testing the truth of a hypothesis.
- A **hypothesis** is a theory or statement whose truth has yet to be proven or disproved.
- The hypothesis we are testing is known as the **null hypothesis** and denoted by H_0.

Hypothesis

- Making a significance test we are testing the truth of a hypothesis.
- A **hypothesis** is a theory or statement whose truth has yet to be proven or disproved.
- The hypothesis we are testing is known as the **null hypothesis** and denoted by H_0.
- In statistical theory the null hypothesis H_0 assumed to be true unless the data indicates otherwise.

Null Hypothesis

- $H_0 : \mu = \mu_0$

Null Hypothesis

- $H_0 : \mu = \mu_0$
- where μ is the unknown but true value of the mean while μ_0 is a known hypothesized value of the true mean.

Null Hypothesis

- $H_0 : \mu = \mu_0$
- where μ is the unknown but true value of the mean while μ_0 is a known hypothesized value of the true mean.
- The measured value (from a sample) is $\overline{x}$ is most likely different from the value of μ stated by the H_0 (i.e. $\overline{x} \neq \mu_0$)

Null Hypothesis

- $H_0 : \mu = \mu_0$
- where μ is the unknown but true value of the mean while μ_0 is a known hypothesized value of the true mean.
- The measured value (from a sample) is $\overline{x}$ is most likely different from the value of μ stated by the H_0 (i.e. $\overline{x} \neq \mu_0$)
- Working under the assumption that H_0 is true, we calculate the probability that the observed difference between the sample statistic $\overline{x}$ and the true value of the parameter μ_0 arises solely as a result of random errors.

- The lower the probability of the observed difference occurs by chance, the less likely is that the H_0 is true.

- The lower the probability of the observed difference occurs by chance, the less likely is that the H_0 is true.
- Usually if the probability of such difference is less than 0.05, we reject H_0.

- The lower the probability of the observed difference occurs by chance, the less likely is that the H_0 is true.
- Usually if the probability of such difference is less than 0.05, we reject H_0. In such cases we say the difference is significant at $\alpha = 0.05$ level.

- The lower the probability of the observed difference occurs by chance, the less likely is that the H_0 is true.
- Usually if the probability of such difference is less than 0.05, we reject H_0. In such cases we say the difference is significant at $\alpha = 0.05$ level.
- If H_0 is rejected, we produce evidence that the alternative hypothesis $H_a : \mu \neq \mu_0$ is true.

In a new method for determining selenourea in water , the following values were obtained for tap water samples spiked with 50ng ml^{-1} of selenourea.
50.4, 50.7, 49.1, 49.0, 51.1
Test $H_0 : \mu = 50$
$H_a : \mu \neq 50$
The sample size n=5.
Calculate the sample mean and standard deviation in *R*.
$> \ sel \ <- c(50.4, 50.7, 49.1, 49.0, 51.1)$
$> \ mean(sel)$
[1] 50.06
$> \ sd(sel)$
[1] 0.9555103
The sample mean is $\overline{x}$= 50.06
The sample standard deviation is s= 0.9555103

- Calculate the standard error of the sample mean $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$

- Calculate the standard error of the sample mean
  $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$
- Calculate the Test Statistic
  t=$\frac{observed\ \ value\ -\ hypothesised\ \ value}{standard\ \ error(observed\ \ value)} = \frac{\overline{x}-\mu_0}{SE(\overline{x})} = \frac{50.06-50}{0.4273172} = 0.14$

- Calculate the standard error of the sample mean
  $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$
- Calculate the Test Statistic
  t=$\frac{observed\ value\ -\ hypothesised\ value}{standard\ error(observed\ value)} = \frac{\overline{x}-\mu_0}{SE(\overline{x})} = \frac{50.06-50}{0.4273172} = 0.14$
  This test statistic has a *t* distribution with *n -1*=4 degrees of freedom.

- Calculate the standard error of the sample mean
  $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$
- Calculate the Test Statistic
  t=$\frac{observed\ \ value\ -\ hypothesised\ \ value}{standard\ \ error(observed\ \ value)} = \frac{\overline{x}-\mu_0}{SE(\overline{x})} = \frac{50.06-50}{0.4273172} = 0.14$
  This test statistic has a *t* distribution with *n -1*=4 degrees of freedom.
- Decide that the significance level is $\alpha = 0.05$.

- Calculate the standard error of the sample mean
  $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$
- Calculate the Test Statistic
  t=$\frac{observed\ \ value\ -\ hypothesised\ \ value}{standard\ \ error(observed\ \ value)} = \frac{\overline{x}-\mu_0}{SE(\overline{x})} = \frac{50.06-50}{0.4273172} = 0.14$
  This test statistic has a *t* distribution with *n -1*=4 degrees of freedom.
- Decide that the significance level is $\alpha = 0.05$.
- The critical value is $t_{\frac{\alpha}{2};n-1} = t_{\frac{0.05}{2};5-1} = t_{0.025;4} = 2.776445$ and can be obtained from *R* with the command
  $> \ qt(.975, 4)$

- Calculate the standard error of the sample mean $SE(\overline{x}) = \frac{s}{\sqrt{n}} = \frac{0.9555103}{\sqrt{5}} = 0.4273172$
- Calculate the Test Statistic t=$\frac{observed\ \ value\ -\ hypothesised\ \ value}{standard\ \ error(observed\ \ value)} = \frac{\overline{x}-\mu_0}{SE(\overline{x})} = \frac{50.06-50}{0.4273172} = 0.14$
  This test statistic has a *t* distribution with *n -1*=4 degrees of freedom.
- Decide that the significance level is $\alpha = 0.05$.
- The critical value is $t_{\frac{\alpha}{2};n-1} = t_{\frac{0.05}{2};5-1} = t_{0.025;4} = 2.776445$ and can be obtained from *R* with the command
  $> \ qt(.975, 4)$

Decision Rule: the Test Statistic t=0.14 and is less than the critical value of 2.776445, hence we fail to reject H_0

Hypothesis testing in R

$> \ sel < -c(50.4,\ 50.7,\ 49.1,\ 49.0,\ 51.1)$

$> \ t.test(sel, mu = 50)$

One Sample t-test

data: sel

t = 0.1404, df = 4, p-value = 0.8951

alternative hypothesis: true mean is not equal to 50

95 percent confidence interval:

48.87358 51.24642

sample estimates:

mean of x

50.06

Decision Rule: The p-value = 0.8951 which is greater than the significance level α=0.05 so we fail to reject the H_0.

Example 3.3.2

In a series of experiments on the determination of tin in foodstuffs, samples were boiled with hydrochloric acid under reflux for two different times: 30 and 75.

refluxing time(min)	Tin found
30	55,57,59,56,56,59
75	57,55,58,59,59,59

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

Test $H_0 : \mu_1 = \mu 2$

$H_a : \mu_1 \neq \mu 2$

Assume the 2 samples have standard deviations are not significantly different.

Two independent samples t-test.

The t.test( ) function produces a variety of t-tests.
When comparing means from two separate populations t.test( ) assumes by default unequal variance.

$> \; x < -c(55, 57, 59, 56, 56, 59)$

$> \; y < -c(57, 55, 58, 59, 59, 59)$

$> \; t.test(x, y, var.equal = TRUE)$

Two Sample t-test

data: x and y t = -0.8811, df = 10, p-value = 0.3989 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval:
-2.940597 1.273931
sample estimates: mean of x mean of y
57.00000 57.83333

Example 3.3.3

The concentration of thiol in the blood lysate in two groups of volunteers, this first group being normal and the second suffering from arthritis:

Normal	1.85	1.92	1.94	1.92	1.85	1.91	2.07
Arthritis	2.81	4.06	3.62	3.27	3.27	3.76	

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

Test $H_0 : \mu_1 = \mu 2$
$H_a : \mu_1 \neq \mu 2$

Assume the 2 samples have significantly different standard deviations.

Two independent samples t-test.

The t.test( ) function produces a variety of t-tests.
When comparing means from two separate populations t.test( ) assumes by default unequal variance.

$> \ x < -c(1.85, 1.92, 1.94, 1.92, 1.85, 1.91, 2.07)$
$> \ y < -c(2.81, 4.06, 3.62, 3.27, 3.27, 3.76)$
$> \ t.test(x, y)$

```
        Welch Two Sample t-test
data: x and y
t = -8.4741, df = 5.241, p-value = 0.0002974
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-2.003538 -1.080748
sample estimates:
mean of x mean of y
1.922857 3.465000
```