22.4UC2 Hypothesis Testing Using MATLAB

MATLAB provides mainly three functions for hypothesis testing, `ttest`, `ttest2`, and `ztest`.

The Z-test

The syntax for this command is

```
h = ztest(x,m,sigma,alpha)
h = ztest(x,m,sigma)
[h,sig,ci] = ztest(x,m,sigma,alpha,tail)
```

The first command performs a two-tailed Z-test to determine whether a sample `x` from a normal distribution with standard deviation `sigma` has mean `m`. The level of significance is `alpha`. The possible outcomes for `h` are `0`, we do not reject the null hypothesis $\bar{x} = m$, or `1`, we do reject the null hypothesis.
The second command does the same as the first, with fixed level of significance `alpha` $= 0.05$.
The third command allows to specify in addition a `tail`, which can take the values 0, -1, or 1:

- `tail = 0` specifies the (default) alternative hypothesis $\bar{x} \neq m$.
- `tail = -1` specifies the alternative hypothesis $\bar{x} < m$.
- `tail = 1` specifies the alternative hypothesis $\bar{x} > m$.

The output of all three commands is by default either 0 or 1, the value of `h`. If we specify `[h,sig,ci]=z(...)` or `[h,sig]=z(...)` or ... then the output is a vector with the following meaning:

- `h` gives the result whether the null hypothesis is to reject or not.
- `sig` is the probability that the observed value of the Z-distribution could be larger by chance (under the assumption of the null hypothesis, i.e., that $\bar{x} = m$).
- `ci` is a $(1 - \alpha)100\%$ confidence interval for the true population mean.

The following example generates 50 random numbers for the standard normal distribution which are stored in the variable `x`. The mean is typed out, as well as `[h,sig,ci]`. In a second test (with the same random numbers) we check the hypothesis that the data have mean 2.

```
>> x = normrnd(0,1,50,1);
 m = mean(x)
>> =
    0.0393
>> [h,sig,ci] = ztest(x,0,1,0.01)
h =
     0
sig =
    0.7809
ci =
   -0.3249    0.4036
>> [h,sig,ci] = ztest(x,2,1,0.05)
h =
     1
```

```
sig =
  1.0464e-043
ci =
   -0.2378    0.3165
>>
```

The t-test

The syntax for the t-test is

```
h = ttest(x,m,sigma,alpha)
h = ttest(x,m,sigma)
[h,sig,ci] = ttest(x,m,sigma,alpha,tail)
```

The t-test is used as the Z-test above, where again `sig` denotes the P-value associated with the t-statistics

$$t = \frac{\bar{x} - m}{s/\sqrt{n}} \, ,$$

where s is the sample standard deviation and n the number of observations.

Testing for the difference between means

The syntax for the t-test testing for the difference between two population means (with unknown but equal standard deviations) from normal populations is

```
h = ttest2(x,y,alpha)
h = ttest2(x,y)
[h,sig,ci] = ttest2(x,y,alpha,tail)
```

Here `x` and `y` are the means of the two samples. The significance `sig` is again the P-value of the associated statistics

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} \, ,$$

where s is the pooled sample standard deviation and n and m are the numbers of observations.