HA10 - Theorectical

February 8, 2018

1 Huajie Chen

2 4199962

```
In [1]: import numpy as np
        import scipy.stats as stats
        import matplotlib as plt
```

3 a.1

3.0.1 Setting up Null-Hypothesis

$$H_0 = \text{The mean of the sample and the given } \mu_0 \text{ are identical.}$$

$$H_1 = \text{The mean of the sample and the given } \mu_0 \text{ are significantly different.}$$

$$\alpha = 0.05$$

3.0.2 Since the variance of the original normal distribution is unknown, and it is now the mean of the sample and the given original mu_0 being compared, the two-side t-test between a specific expectation and a group of data should be performed.

$$\bar{x} = \frac{1}{k} \sum_{k=1}^{k} x_k$$

$$x_{var} = \frac{\sum_{k=1}^{k} (x_k - \bar{x})^2}{k}$$

$$t_{value} = \frac{\bar{x} - \mu_0}{x_{var}}$$

3.0.3 Calculating the confidence interval

$$CI = (norm.ppf(\alpha/2),\ norm.ppf(1 - \alpha/2))$$

(norm.ppf is the percent point function in python)

3.0.4 If the t-value is located in this interval, the Null-Hypotheis will not be rejected. If not, it should be rejected.

4 a.2

4.0.1 This test may be applied to i.e. the expression data of normal liver cell and liver cancer cell. If the test of one specific gene in the normal cell differs from the other, we may tend to believe that there is research value underneath.

5 b.1

```
In [7]:  # b.1
         mu = 7.3
         sigma = 3.2
         alpha1 = 0.05
         alpha2 = 0.01

         dist = stats.norm(mu, sigma)

         print('The expression range containing values with 95% should be
               ({:.3f}, {:.3f})'.format(dist.ppf(alpha1/2), dist.ppf(1-alpha1/2)))
         print('The expression range containing values with 99% should be
               ({:.3f}, {:.3f})'.format(dist.ppf(alpha2/2), dist.ppf(1-alpha2/2)))
```

```
The expression range containing values with 95% should be (1.028, 13.572)
The expression range containing values with 99% should be (-0.943, 15.543)
```

6 b.2

```
In [8]:  # b.2
         print('The expression range for the first quartile should be
               (-infinity, {:.3f})'.format(dist.ppf(0.25)))
```

```
The expression range for the first quartile should be (-infinity, 5.142)
```

7 b.3

```
In [11]:  # b.3
          print('The probability of a gene having an expression value above 10 should be
                {:.3f}'.format(1-dist.cdf(10)))
```

```
The probability of a gene having an expression value above 10 should be 0.199
```