Tai Duc Nguyen - ECEC 487 - 11/03/2019

Homework 6

Programming Homework

```
clear all; close all;
randn('seed', 0);
```

Problem 5.1a Textbook p.316

```
mu1 = 0; sigma1 = 1; n1 = 100;
N1 = normrnd(mu1, sigma1, [n1 1]);

mu2 = 2; sigma2 = 1; n2 = 100;
N2 = normrnd(mu2, sigma2, [n2 1]);

[H, rho] = ttest2(N1, N2);
sprintf("For N1=N(%.2f,%.2f), %d samples and N2=N(%.2f,%.2f), %d samples
    mu1, sigma1, n1, mu2, sigma2, n2, H, rho)
```

```
ans =
"For N1=N(0.00,1.00), 100 samples and N2=N(2.00,1.00), 100
samples, H and P are: 1, 0.0000"
```

Problem 5.1b Textbook p.316

```
mu1 = 0; sigma1 = 1; n1 = 100;
N1 = normrnd(mu1, sigma1, [n1 1]);
```

```
mu2 = 0.2; sigma2 = 1; n2 = 100;
N2 = normrnd(mu2, sigma2, [n2 1]);

[H, rho] = ttest2(N1, N2);
sprintf("For N1=N(%.2f,%.2f), %d samples and N2=N(%.2f,%.2f), %d samples
    mu1, sigma1, n1, mu2, sigma2, n2, H, rho)
```

```
ans =
"For N1=N(0.00,1.00), 100 samples and N2=N(0.20,1.00), 100
samples, H and P are: 0, 0.1604"
```

Problem 5.1c Textbook p.316 part 1

```
mu1 = 0; sigma1 = 1; n1 = 150;
N1 = normrnd(mu1, sigma1, [n1 1]);

mu2 = 2; sigma2 = 1; n2 = 200;
N2 = normrnd(mu2, sigma2, [n2 1]);

[H, rho] = ttest2(N1, N2);
sprintf("For N1=N(%.2f,%.2f), %d samples and N2=N(%.2f,%.2f), %d samples
    mu1, sigma1, n1, mu2, sigma2, n2, H, rho)
```

```
ans =
"For N1=N(0.00,1.00), 150 samples and N2=N(2.00,1.00), 200
samples, H and P are: 1, 0.0000"
```

Problem 5.1c Textbook p.316 part 2

```
mu1 = 0; sigma1 = 1; n1 = 150;
N1 = normrnd(mu1, sigma1, [n1 1]);

mu2 = 0.2; sigma2 = 1; n2 = 400;
N2 = normrnd(mu2, sigma2, [n2 1]);

[H, rho] = ttest2(N1, N2);
sprintf("For N1=N(%.2f,%.2f), %d samples and N2=N(%.2f,%.2f), %d samples
    mu1, sigma1, n1, mu2, sigma2, n2, H, rho)
```

```
ans =
"For N1=N(0.00,1.00), 150 samples and N2=N(0.20,1.00), 400
samples, H and P are: 0, 0.1029"
```

Problem 5.1c Textbook p.316 verification

```
mu1 = 0; sigma1 = 1; n1 = 150;
N1 = normrnd(mu1, sigma1, [n1 1]);

mu2 = 0.2; sigma2 = 1; n2 = 400;
N2 = normrnd(mu2, sigma2, [n2 1]);

[H, rho] = ttest2(N1, N2);
sprintf("For N1=N(%.2f,%.2f), %d samples and N2=N(%.2f,%.2f), %d samples
    mu1, sigma1, n1, mu2, sigma2, n2, H, rho)
```

```
ans =
"For N1=N(0.00,1.00), 150 samples and N2=N(0.20,1.00), 400
samples, H and P are: 1, 0.0038"
```

Conclusion for Problem 5.1

In this problem, MATLAB's `ttest2` function, which performs t-test on 2 samples (assuming they come from the Normal Distribution with the null hypothesis that the they have equal means: `H=0` means that the null hypothesis "cannot be rejected at the 5% signifiance level". `H=1` means that the null hypothesis "can be rejected at the 5% level". For part A, since the means of the 2 samples are far apart (0 and 2), H=1 -- the 2 data sets have significantly different means. when the 2 means are very close (0 and 0.2) in part B, the t-test results in `H=0` -- the 2 data sets' means are not significantly different. In part C, the same results come out of the t-test for the 2 cases. However, in the case where the means are very close, as the number of points in data set 2 increases and that of data set 1 stays the same, ρ ("the probability of observing the given result, or one more extreme, by chance if the null hypothesis is true") decreases and `H=1` (shown in the verification part of 5.1c). This observation is consistent to the intuition that the more points in a data set, the better the chances that the mean of such data set equal to the *TRUE* mean.

Written Homework

Answering question on how the table 5.1 in Textbook p.271 got generated

Since the t-test statistic is:

$$q = \frac{\text{sample mean} - \text{given mean}}{\text{sample variance}/\sqrt(N)}$$

, a distribution that's made up of this statistical variable can be made by considering the Central Limit Theorem -- $N(\text{given mean}, \text{sample variance}^2/N)$. When N tends to infinity, this distribution tends toward the Normal Distribution with 0 mean and unit variance. Hence, the acceptance interval in table 5.1 can be calculated using $N(0, 1)$ and $1 - \rho$ (a.k.a the ratio of the area under the curve vs the area of the entire distribution) => The 95 confidence interval -- or, the 5 acceptance interval is 1.967.