Exercise 4

Part 1

1. Open the file normtemp.mat that contains the measurements of normal body temperature taken on healthy people. The three columns in the dataset are temperature, gender, and heart rate. We will need only the first column.

2. If you prefer working with degrees Celsius, you do the conversion:

   tempC = (normtemp(:,1) - 32) * 5/9

3. Plot the histogram. Contemplate if the data is normally-distributed. Note the total number of measurements N.

4. Calculate the mean and the variance of the distribution. Remember to use the formula for variance that uses (N-1), rather than N. Does the calculated mean correspond to the value that you thought was supposed to be the normal body temperature? (The "supposed normal" actually may depend on the country where you were born! I personally always thought it was 36.6 C = 97.88 F.) We are going to calculate the probability that your calculated mean is consistent with the "supposed normal" temperature.

5. Calculate the uncertainty on the mean, s_{μ}.

6. Now, calculate the probability that your calculated mean is consistent with the "supposed normal" temperature. First, calculate how many standard deviations (s_{μ}, not σ!) your mean value is away from the "supposed normal". For me, this would be something like t=abs((meanC-36.6)/smu), where variables meanC and smu are the mean and the error on the mean. This gives the value of *t* in the Student's distribution, i.e. how far away you are from the "true" mean, in units of s_{μ}.

7. Using t-distribution, calculate the probability of being this far away from the "true" mean. You will probably want to use the tcdf(t,N-1) function.

8. Verify your result by calculating the 99% confidence interval for the mean (i.e., the temperature range where the true mean should fall). You can use either *t*- or Gaussian distributions. They should be similar for this large N.

Additional question 1: When would you consider someone's temperature abnormal? I.e., when should the physician give you a sick leave from the class?

Additional question 2: Is there a significant difference between the males and females in terms of the normal body temperature? In the second column of your dataset, 1=male, 2=female.

Additional question 3: You may have noticed that there are several temperature measurements that appear abnormally high or abnormally low for a healthy person. Is it possible that those persons were actually sick at the time of measurements? In other words, should you exclude those "abnormal" measurements from the statistics?

Exercise 4, Part 2:

9. We now want to investigate how the total number of measurements affects the confidence of your conclusions. Suppose that instead of 130 measurements, you only made 6. To simulate this, let's randomly pick 6 values from the dataset:

```
tempR(1:6)=0;
tempR(1)=tempC(floor(unifrnd(1,131)))
tempR(2)=tempC(floor(unifrnd(1,131)))
tempR(3)=tempC(floor(unifrnd(1,131)))
tempR(4)=tempC(floor(unifrnd(1,131)))
tempR(5)=tempC(floor(unifrnd(1,131)))
tempR(6)=tempC(floor(unifrnd(1,131)))
```

   This stores 6 randomly chosen values in the variable tempR. Plot them in a histogram.

10. Calculate the mean, variance, and the error on the mean.

11. With such a small sample, it is difficult to predict where your mean will be relative to the "true" mean, so instead of calculating the *t* value, let's use the t-distribution to calculate the 99% confidence interval. Table C.8 in Bevington's book tells me that, for N=6, the 99% confidence corresponds to exactly 4*t*. (Which value did I use for ν?) You can verify this result using tcdf. So, for your 6-point dataset, what are the minimum and maximum temperatures that correspond to this confidence interval? Is this interval for the value of the mean smaller or larger than the spread of your measured temperature values in the histogram?

The 95% confidence for N=6 approximately corresponds to t=2.6. See how this changes your result.

12. Compare the confidence limits that you obtained using Student's *t* to what you'd have obtained using Gaussian. Gaussian probability 99% corresponds to 2.6*t* (or $2.6\sigma_\mu$), whereas probability 95% approximately corresponds to 2*t* (regardless of the number of measurements!)