ECE 376 - Homework #6

Statistics and Data Collection. Due Monday, October 11th

Chi-Squred Test

The following code implements a fair die and a loaded die (with the comment removed).

```
while(1) {
   while(!RB0);
   while(RB0) {
      d7 = (d7 + 1) % 7;
      d100 = (d100 + 1) % 100;
      }
    d7 = d7 + 1;
// Loaded Die
// if(d100 < 10) d7 = 7;
    LCD_Move(1,8); LCD_Out(d7, 1, 0);
    SCI_Out(d7, 1, 0);
    SCI_CRLF();
    }
```

1) Collect data for the fair 7-sided die. From your data, what is the probability that the die is fair?

Raw Data: 185 die rolls

1	2	3	4	5	6	7
21	24	14	29	30	32	35

Setting up a table to compute the Chi-Sqquared score

Die Roll	p()	expected	actual	chi-squared
1	1/7	26.43	21	1.12
2	1/7	26.43	24	0.22
3	1/7	26.43	14	5.85
4	1/7	26.43	29	0.25
5	1/7	26.43	30	0.48
6	1/7	26.43	32	1.17
7	1/7	26.43	35	2.78
				11.87

From StatTrek,

- A chi-squared score of 11.87
- With 6 degrees of freedom
- Has a probability of 0.94

From this data, there is a 94% chance that the die is not fair

2) Remove the comment and collect data for the loaded die. From your data, what is the probaility that the die is fair?

Raw Data: 218 die rolls

1	2	3	4	5	6	7
26	38	27	25	32	20	50

Setting up a chi-squared table

Die Roll	p()	expected	actual	chi-squared
1	1/7	31.34	26	0.91
2	1/7	31.34	38	1.42
3	1/7	31.34	27	0.6
4	1/7	31.34	25	1.28
5	1/7	31.34	32	0.01
6	1/7	31.34	20	4.1
7	1/7	31.34	50	11.11
			Total	**19.44**

From StatTrek, a chi-squared score of 19.44 corresponds to a probability of 0.997

I am 99.7% certain that this is a loaded die

3) How loaded does the die have to be for you to be able to reliably detect that something is amiss?

Assume

- 210 die rolls
- 95% certain that the die is loaded
- X too many 7's
- Each other number is X/6 too few

From Stat-Trek, 95% certainty means the chi-squared score should be 12.6

Die Roll	p()	expected	actual	chi-squared
1	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
2	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
3	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
4	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
5	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
6	1/7	30	30 - X/6	$\left(\frac{X^2}{1080}\right)$
7	1/7	30	30 + X	$\left(\frac{X^2}{30}\right)$
			Total	0.03889 X^2

$$0.03889X^2 = 12.6$$

$$X = 18.00$$

I can load the die so that I get 18 more 7's than I should and get away with it 95% of the time. 18/200 corresponds to 9%

$$loading = \left(\frac{\text{18 extra 7s}}{\text{200 rolls}}\right) = 9\%$$

t-Test

4) Using your data from problem #1 (fair die), determine the 90% confidence interval for any given roll (individual).

```
>> x = mean(D1)
x =    4.4000

>> s = std(D1)
s =    2.0087
```

90% confidence interval with 186 degrees of freedom means

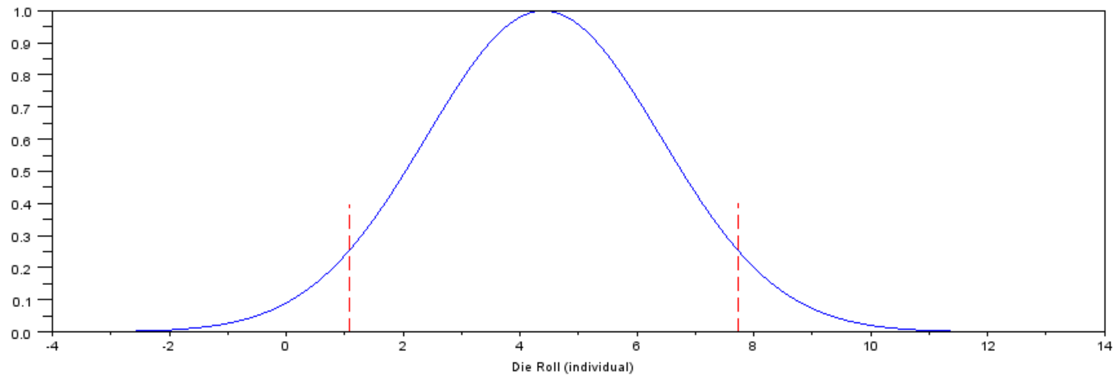
- 5% tails (two tails, each 5%).
- t score = 1.653

$$x - 1.653s < roll < x + 1.653s$$

$$4.40 - 1.653 \cdot 2.0087 < roll < 4.40 + 1.653 \cdot 2.0087$$

$$1.0797 < roll < 7.7203$$ with a probability of 0.9

```
>> x = mean(D1)
x =    4.4000

>> s = std(D1)
s =    2.0087

>> x - 1.653*s
ans =     1.0797

>> x + 1.653*s
ans =     7.7203
```



90% confidence interval for any given die roll

5) Using your data from problem #1 (fair die), determine the 90% confidence interval for the mean of this die (population). From this, is it a fair die?

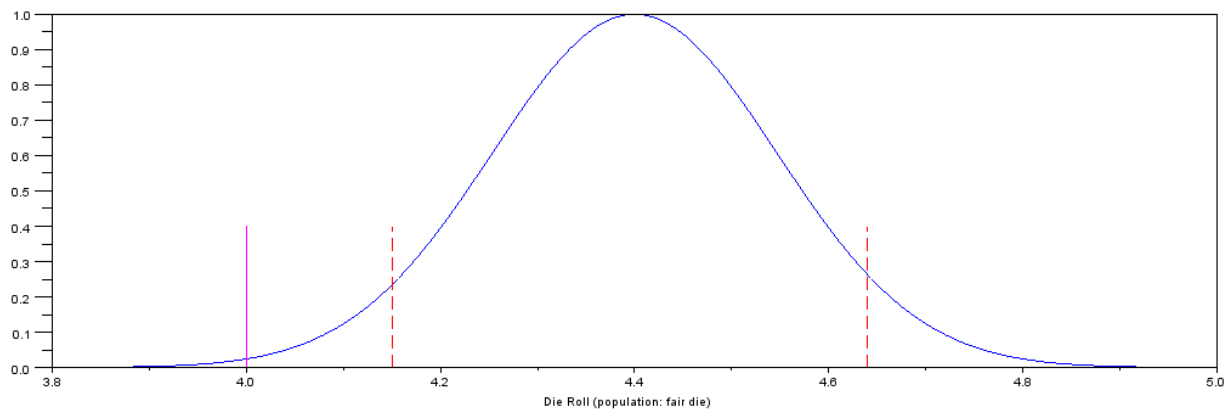
With populations, the standard deviation drops as the square root of the sample size (you know more about the population with more data).

Scaling the standard deviation by $\sqrt{187}$:

```
>> x - 1.653*(s/sqrt(187))

ans =     4.1572

>> x + 1.653*(s/sqrt(187))

ans =     4.6428
```

I am 90% certain that the die's mean is somewhere in the range of (4.1572, 4.6428).

- Since 4.000 (a fair die) is not in this range, I conclude the die is not fair



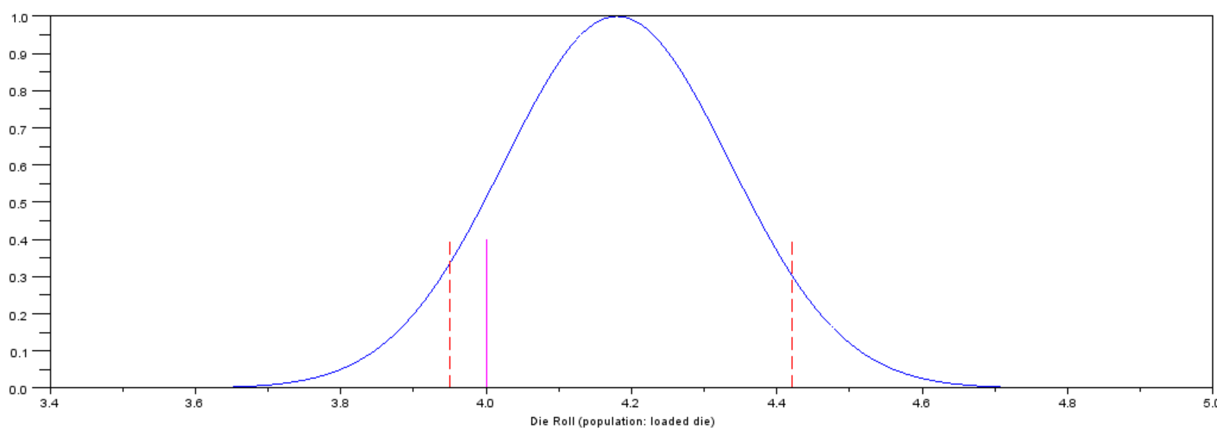
90% Confidence Interval for the mean of a fair die (population)
Since 4.000 isn't in this interval, I can be 90% certain that this die is not fair
(i.e. the mean is not 4.00)

6) Using your data from problem #2 (loaded die), determine the 90% confidence interval for the mean of this die (population). From this, is it a fair die?

```
>> x = mean(D2)
x =    4.1881

>> s = std(D2)
s =    2.1113

>> x - 1.653*(s/sqrt(218))
ans =    3.9517

>> x + 1.653*(s/sqrt(218))
ans =    4.4244
```

I am 90% certain that the die's mean is somewhere in the range of (3.9517, 4.4244).

- Since 4.000 (a fair die) is in this range, I conclude the die is not loaded



90% confidence interval for the mean of a loaded die (population)
Since 4.00 is within this interval, I can be 90% certain that this is a fair die

7) How many times would you need to roll the die to detect

- A shift in the mean of 0.2 or more
- With 90% certainty?

Assume

- mean = 4.2
- standard deviation is 2.000

$$1.653 = \left(\frac{0.2}{\left(\frac{2}{\sqrt{n}} \right)} \right) = 0.1\sqrt{n}$$

n=273

If your sample size is 273, you should be able to detect loading which shifts the mean by 0.2/

Data Collection & Analysis

8) Measure the resistance of 10 resistors of the same value (100 Ohm, 2.2k Ohm, your pick). From this data, determine

- The 90% confidence interval for any given resistor (individual).
- The probability that a given resistor will be out of tolerance (difffer from marked value by more than 1%, 2%, 5%)

raw data: 200 Ohm 5% resistors:

195.2, 195.6, 197.4 197.2 196.0 197.7 196.4 194.8 197.2 197.9

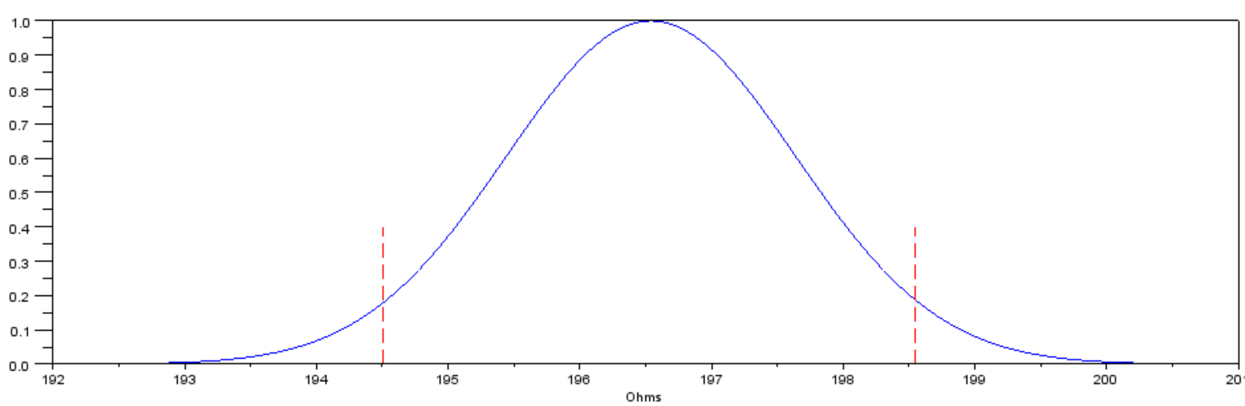
```
>> R = [195.2,  195.6, 197.4 197.2  196.0  197.7  196.4  194.8  197.2  197.9];
>> x = mean(R)
x =  196.5400

>> s = std(R)
s =    1.0967

>> s1 = [-4:0.01:4]';
>> p = exp(-s1.^2 / 2);
>> plot(s1*s+x,p)
>> xlabel('Ohms');
>> ylabel('Probability')
```

90% confidence interval for any given resistor:

- t-score for 5% tails with 9 degrees of freedom is 1.833

```
>> x - 1.833*s
ans =  194.5298

>> x + 1.833*s
ans =  198.5502
```

The 90% confidence interval for any given 200-ohm resistor is (194.5298, 198.5502) Ohms



90% confidence interval for a 200 Ohm resistor

What is the probability that a given resistor is more than 5% out of tolerance?

Translation: What is

- the probability that a given resistor is less than 190 Ohms (5% low), plus
- the probability that a given resistor is more than 210 Ohms (5% high)?

5% Low: Compute the t-score and find the area of the tail

```
-->t = (x - 190) / s

 t  =    5.9633446
```

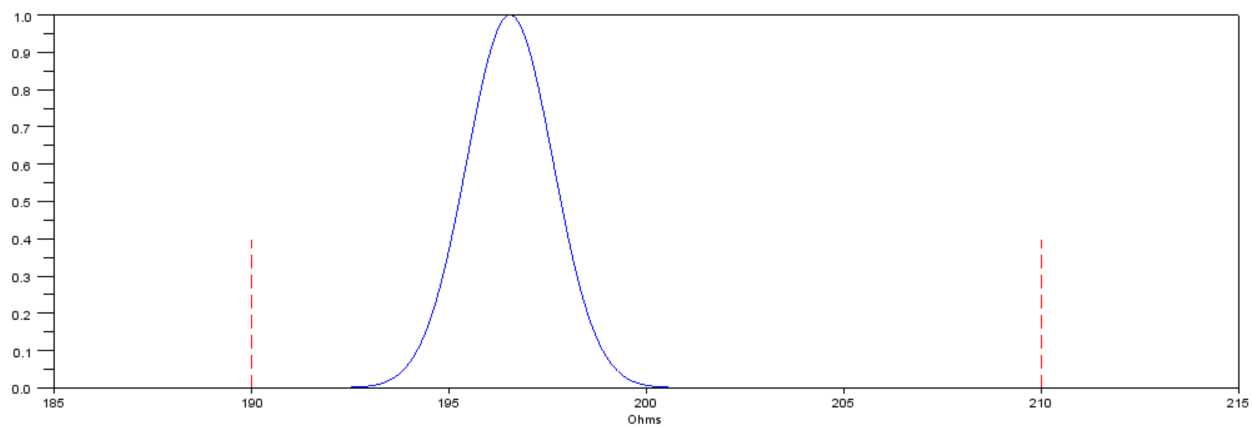
From StatTrek, this corresponds to a probability of 0.0001

5% High: Compute the t-score and find the area of the tail

```
-->t = (210 - x) / s

 t  =    12.273183
```

From StatTrek, this corresponds to a probability of 0.0000

There probability of a resistor being more than 5% out of tolerance is 0.0001



probabilty density function for a 200 Ohm resistor and +/- 5% tolerance bands
p(> 5% out of tolerance) = 0.0001

What is the probability that a given resistor is more than 2% out of tolerance?

Translation: What is

- the probability that a given resistor is less than 196 Ohms (2% low), plus
- the probability that a given resistor is more than 204 Ohms (2% high)?

2% Low: Compute the t-score and find the area of the tail

```
-->t = (x - 196) / s

 t  =    0.4923862
```

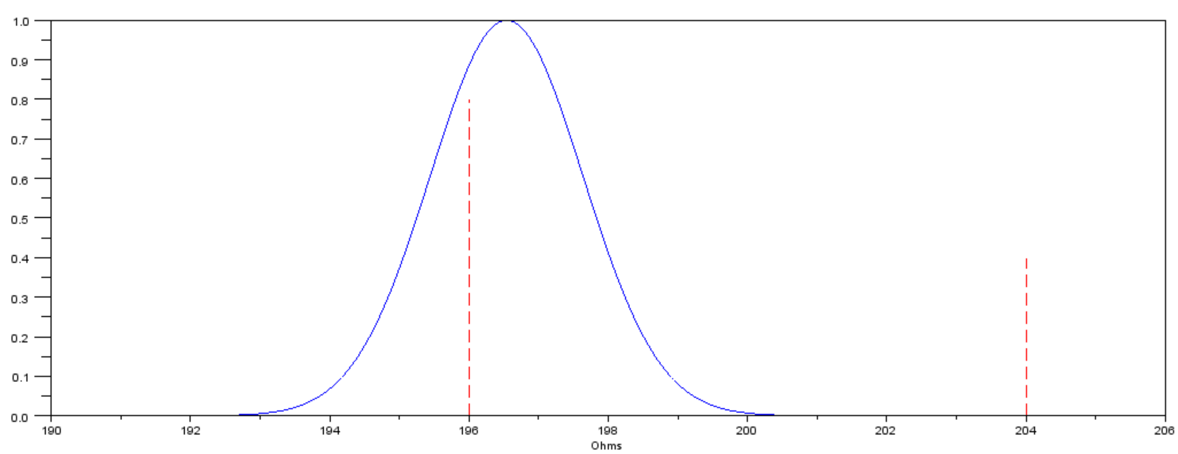
From StatTrek, this corresponds to a probability of 0.3171

5% High: Compute the t-score and find the area of the tail

```
-->t = (204 - x) / s

 t  =    6.8022249
```

From StatTrek, this corresponds to a probability of 0.0000

There probability of a resistor being more than 2% out of tolerance is 0.3171



pdf for a 200 Ohm resistor and +/- 2% Tolerance Bands
p(>2% out of tolerance) = 0.3171

What is the probability that a given resistor is more than 1% out of tolerance?

Translation: What is

- the probability that a given resistor is less than 198 Ohms (1% low), plus
- the probability that a given resistor is more than 202 Ohms (1% high)?

1% Low: Compute the t-score and find the area of the tail

```
-->t = (198 - x) / s

 t  =    1.3312665
```

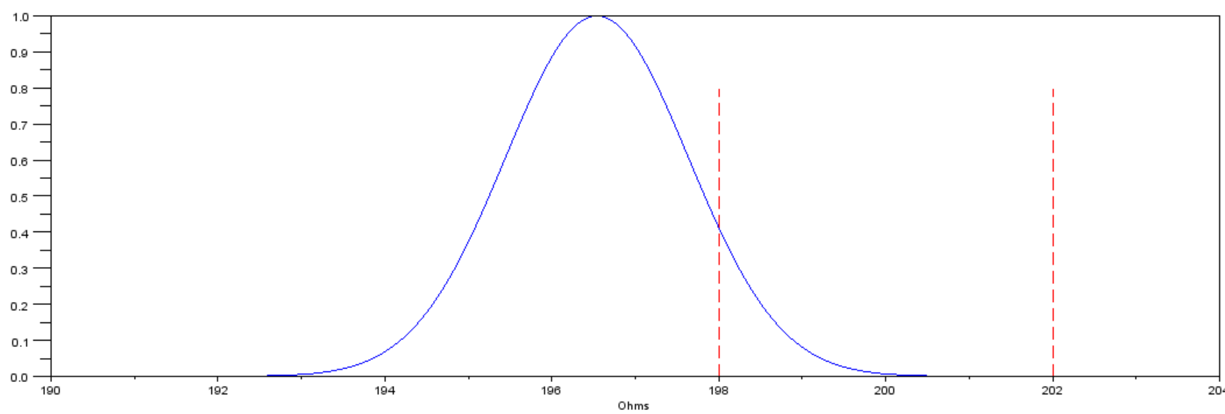
From StatTrek, this corresponds to a probability of 0.8919

1% High: Compute the t-score and find the area of the tail

```
-->t = (x - 202) / s

 t  =  - 4.9785721
```

From StatTrek, this corresponds to a probability of 0.0004

There probability of a resistor being more than 1% out of tolerance is 0.8923



pdf for a 200 Ohm resistor and +/- 1% Tolerance Band
p(>1% out of tolerance) = 0.8923