ECE 111 - Homework #12

Week #12: ECE 341 Random Processes. Due 11am November 15th

Chi-Squared Tests

Problem 1: The following Matlab code generates 240 random die rolls for a six sided die

```
RESULT = zeros(1,6);
for i=1:240
   D6 = ceil( 6*rand );
   RESULT(D6) = RESULT(D6) + 1;
   end
RESULT
```

Determine whether this is a fair or loaded die using a Chi-Squared test.

The results I got were:

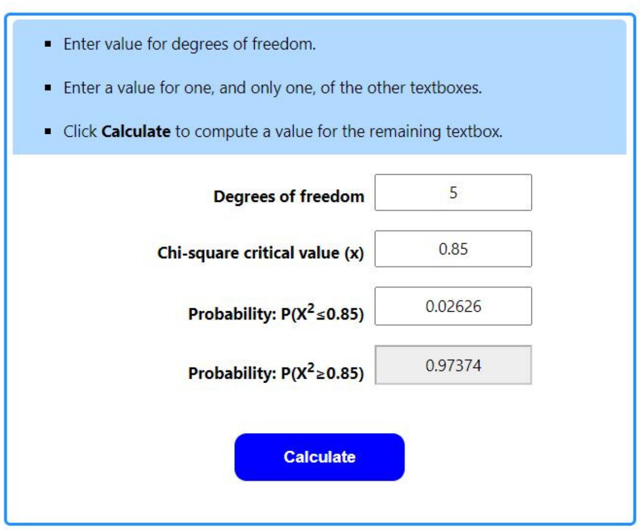
```
RESULT =    40    44    42    38    39    37
```

Calculate the chi-squared score

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	40	40	0
2	1/6	40	44	0.4
3	1/6	40	42	0.1
4	1/6	40	38	0.1
5	1/6	40	39	0.03
6	1/6	40	37	0.23
			Total	0.85

From StatTrek, a chi-squared critical value of 0.85 corresponds to a probability of 0.02626

There is a 2.6% chance this die is loaded



Problem 2: The following Matlab code generates 240 rolls of a loaded six-sided die (5% of the time, you roll a 6):

```
RESULT = zeros(1,6);
for i=1:240
   if(rand < 0.05)
      D6 = 6;
   else
      D6 = ceil( 6*rand );
      end
   RESULT(D6) = RESULT(D6) + 1;
   end
RESULT
```

Determine whether this is a fair or loaded die using a Chi-Squared test.

The result I got was

```
RESULT =  39     30     37     40     42     52
```

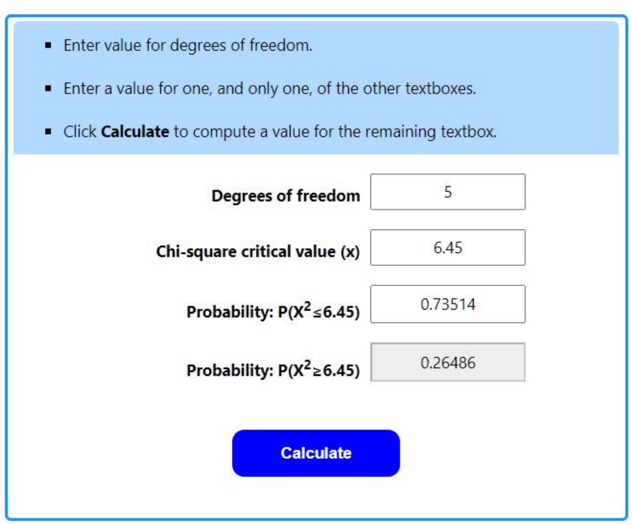
Calculating the Chi-Squared critical value:

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	40	39	0.03
2	1/6	40	30	2.5
3	1/6	40	37	0.23
4	1/6	40	40	0
5	1/6	40	42	0.1
6	1/6	40	52	3.6
			Total	**6.45**

From StatTrek, a Chi-Squred critival value of 6.45 corresponds to a probability of 0.73514

There is a 73.5% chance that this die is loaded

(note: 5% loading is pretty hard to detect)



Am I Psychic?

Problem #3: Shuffle a deck of 52 playing cards and place it face down on a table.

- Predict the suit of the top card then reveal it. If correct, place the card in one pile (correct). If incorrect, place it in another pile.
- Repeat for all 52 cards.

Use a chi-squared test to test the hypothesis that you're just guessing (probability of being correct is 25%)

Pediction	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
Correct	1/4	13	19	2.77
Incorrect	3/4	39	33	0.92
			Total	3.69

Flipping throgh a deck of cards and predicting the suit, I was

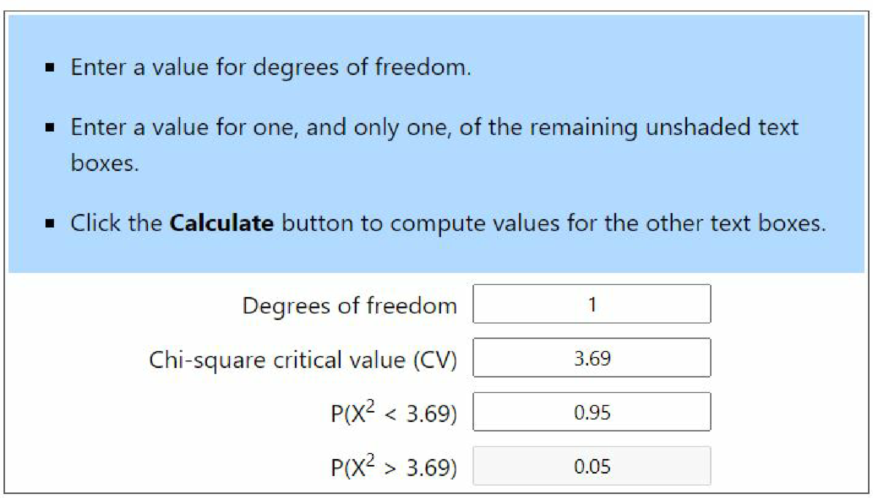
- Correct 19 times
- Incorrect 33 times

Put this data into a table and compute the chi-squared score

From StatTrek, a chi-squared score of 3.69 with 1 degree of freedom corresponds to a probabiliy of 0.95

There is 95% chance that I wasn't just guessing

and a 5% chance I got lucky... before I mortgage the house and go to the cassino, I might want to repeat this test to see if the result is repeatable



Monte-Carlo: y = 2d4 + 3d6 + 4d8

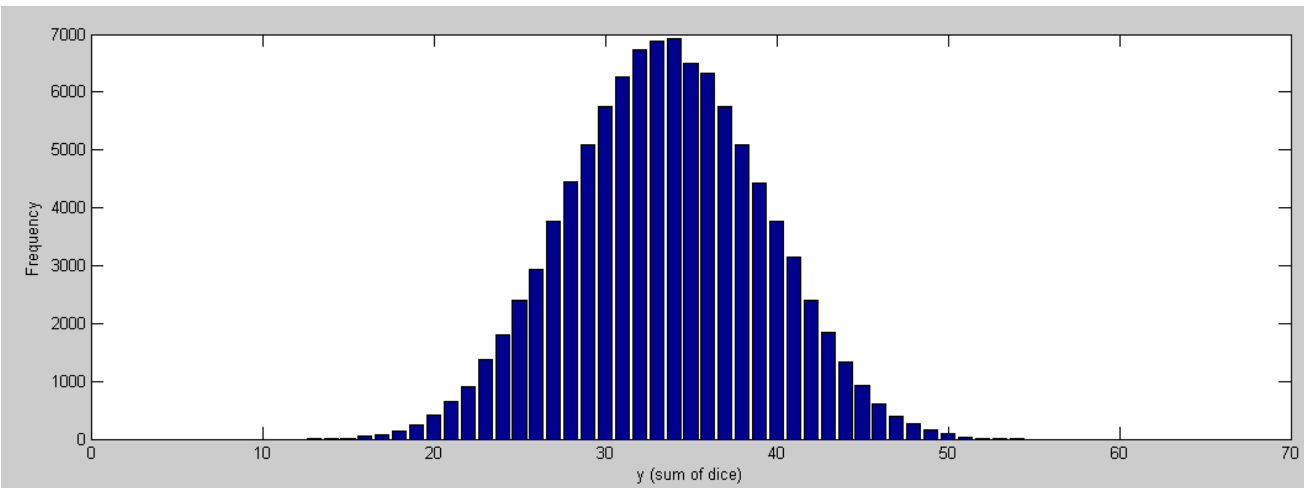
5) Using a Monte Carlo simulation with 100,000 dice rolls, determine

- The probability of rolling 40 or more (y > 39.5)
- The 90% confidence interval for y (5% of the rolls will be less than the lower bound and 5% of the rolls will be more than the upper bound)

Step 1: Roll the dice 100,000 times

- Note that the bar chart is a bell-shaped curve. This is the central limit theorem in action...

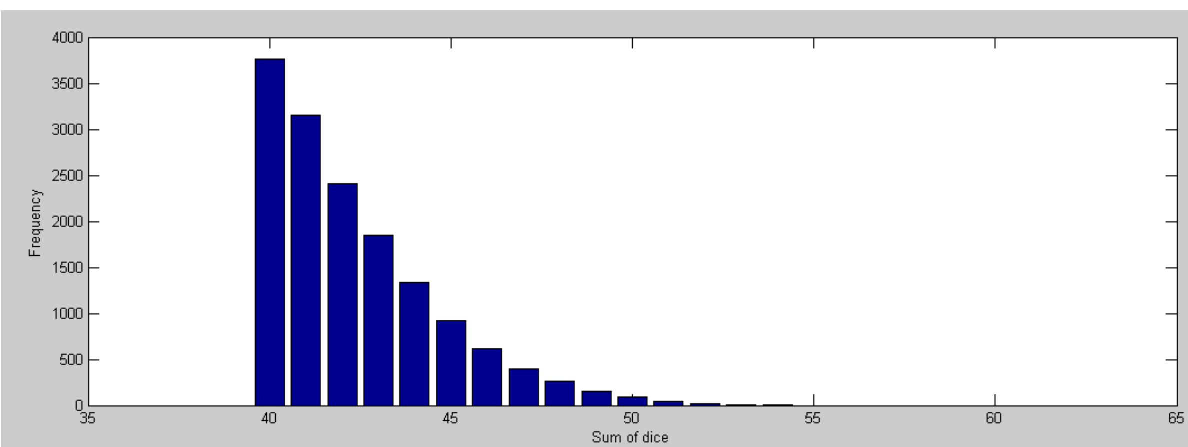
```
RESULT = zeros(1,60);
for n=1:1e5
   d4 = ceil(4*rand(1,2));
   d6 = ceil(6*rand(1,3));
   d8 = ceil(8*rand(1,4));
   y = sum(d4) + sum(d6) + sum(d8);
   RESULT(y) = RESULT(y) + 1;
end
bar(RESULT)
```



a) The probability of rolling 40 or more

Take the data for y is 50 or more:

```
>> bar(RESULT(40:60))
```



Add up the number of times you rolled 40 or more, divided by the sample size (100,000)

```
>> sum(RESULT(40:60)) / 1e5

ans =    0.1503
```

In 100,000 rolls, 15.03% resutled in a sum of 40 or more
- **There is a 15.03% chance of rolling 40 or more**

b) 90% confidence interval: 24 <= roll <= 43
- 90% of the time you will roll numbers bewteen 24 and 43

Upper bound *keep guessing the upper bound until 5% of the results are in the tail*

```
>> sum(RESULT(40:60))/1e5
ans =   0.1503

>> sum(RESULT(41:60))/1e5
ans =    0.1126

>> sum(RESULT(42:60))/1e5
ans =    0.0811

>> sum(RESULT(43:60))/1e5
ans =    0.0570

>> sum(RESULT(44:60))/1e5
ans =    0.0385
```

Lower Bound *keep guessing the upper bound until 5% of the results are in the tail*

```
>> sum(RESULT(1:23))/1e5
ans =    0.0386

>> sum(RESULT(1:24))/1e5
ans =    0.0567

>> sum(RESULT(1:25))/1e5
ans =    0.0807
```

>>

Normal Approximation

Rather than roll the dice 100,000 times, can you compute

- The probablity of rolling more than 39.5, and
- The 90% confidence interval ?

First, determine the mean and standard deviation for a single die

```
d4 = [1,2,3,4];
m4 = mean(d4);                    % mean
v4 = sum( (d4 - m4).^2) / 4;      % variance

d6 = [1,2,3,4,5,6];
m6 = mean(d6);
v6 = sum( (d6 - m6).^2) / 6;

d8 = [1,2,3,4,5,6,7,8];
m8 = mean(d8);
v8 = sum( (d8 - m8).^2) / 8;
```

When you add distributions,

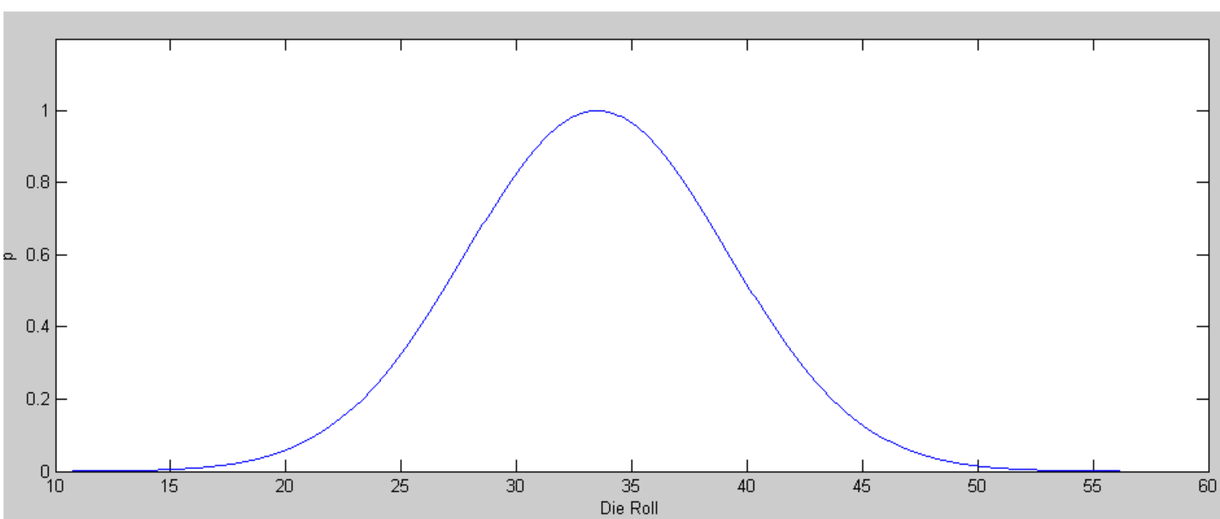
- The means add, and
- The variance adds

```
my = 2*m4 + 3*m6 + 4*m8;    % mean
vy = 2*v4 + 3*v6 + 4*v8;    % variance
sy = sqrt(vy);              % standard deviation

my =   33.5000   mean of y
sy =    5.6789   standard deviation of y
```

You can get an idea of what the distribution looks like using a normal pdf (not required)

```
>> s = [-4:0.01:4]';
>> p = exp(-s.^2 / 2);
>> plot(s*5.6789+33.5,p);
>> xlabel('Die Roll');
>> ylabel('p')
```



Probability y > 39.5

To find the probability of rolling more than 39.5, determine the area to the right (find the z-score)

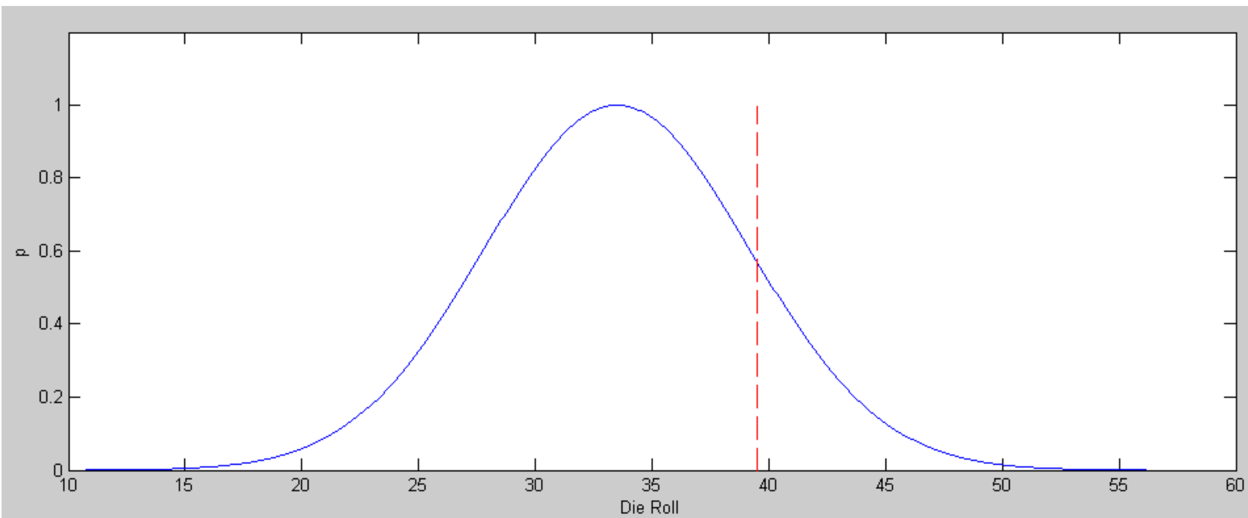
```
>> z = (39.5 - my) / sy
z =    1.0565
```

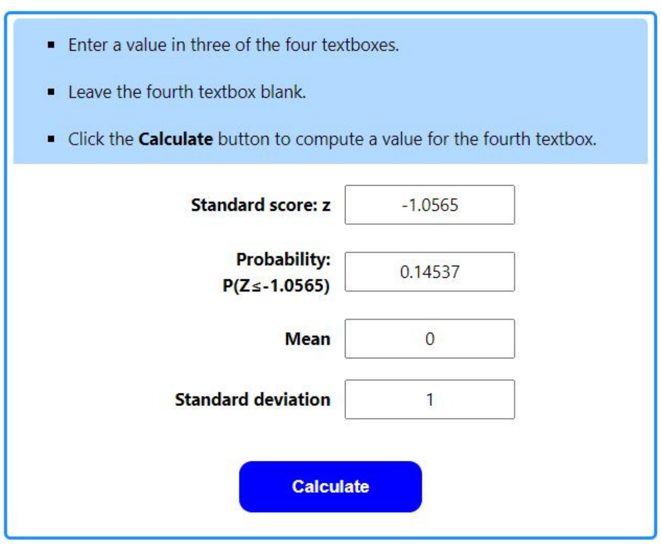
From a normal table (or StatTrek), convert this to a probability

- From StatTrek, this corresponds to a probability of 0.14537
- **There is a 14.537% chance the sum will be more than 39.5**

Note:

- This is almost the same answer we got with 100,000 die rolls with a Monte Carlo simulation
- Zero die rolls were needed to determine this probability
- If it costs $10/roll, that's a lot of money





90% Confidence Interval:

From StatTrek, determine the z-score corresponding to 5% tails

$z = 1.64485$

The 90% confidence interval is then

```
>> Lower = my - 1.64485*sy
>> Upper = my + 1.64485*sy

Lower =   24.1590
Upper =   42.8410
```

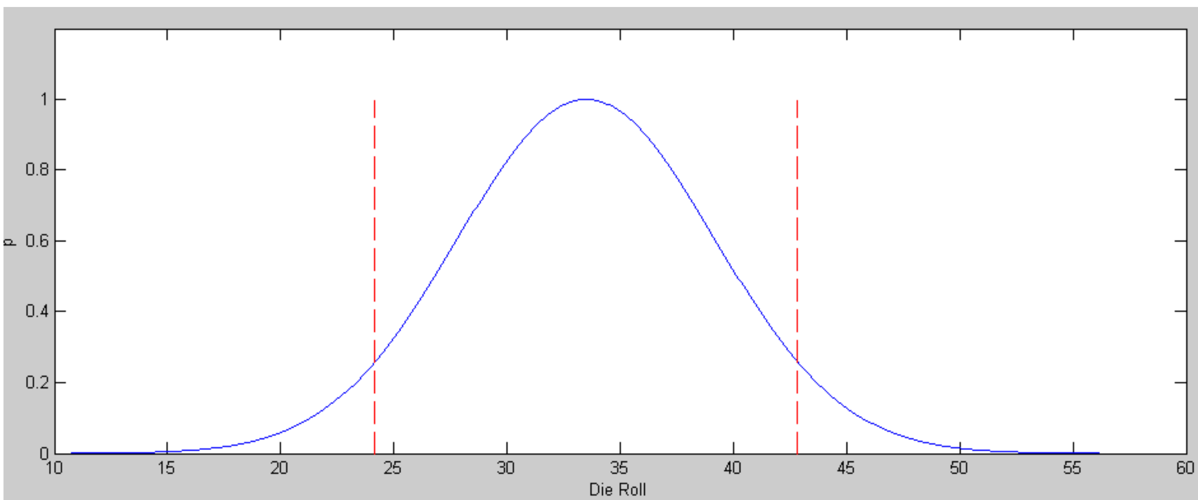
or

$$24.15 < roll < 42.84$$

Note: With a Monte Carlo simulation and 100,000 rolls, the result was

$$24 \le roll \le 43$$

I got this answer using a Normal approximation without having to roll any dice



t-Tests

Suppose you don't know the mean and standard deviation. Can I determine

- The probability of rolling more than 39.5, or
- The 90% confidence interval

without having to roll the dice 100,000 times?

The answer is yes:

- Roll the dice a few times (more than one, less than a million)
- Determine the mean and standard deviation of the result,
- Then use a student-t table to compute these probabilities

Problem 6: Using Matlab, determine five values for Y

$$Y = 2d4 + 3d6 + 4d8$$

Step #1: Collect Data (roll the dice five times)

```
DATA = [];
for i=1:5
   d4 = ceil( 4*rand(2,1) );
   d6 = ceil( 6*rand(3,1) );
   d8 = ceil( 8*rand(4,1) );
   Y = sum(d4) + sum(d6) + sum(d8);
   DATA = [DATA, Y];
   end

DATA =    32    39    38    35    41
```

Step 2: Calculate the mean and standard deviation from your data

```
x = mean(DATA)
s = std(DATA)
n = length(DATA)

x =    37              mean
s =    3.5355          standard deviation
n =    5               sample size
```

Step 3: Use a student-t test to answer your questions

What is the probabilitu of rolling more than 39.5?

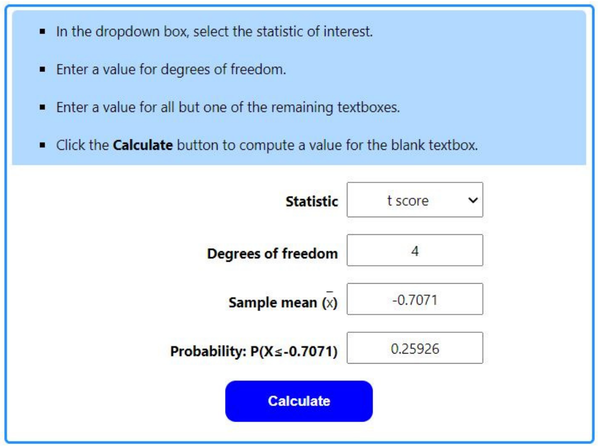
Use a t-test to determine the probabillity of scoring more than 39.5 points. The t-score is

```
>> t = (39.5 - x) / s

t =     0.7071
```

From StatTrek, this corresponds to p = 0.2592

There is a 25.92% chance the sum will be more than 39.5



What is the 90% confidence interval?

From StatTrek, 5% tails along with 4 degrees of freedom corresponds to a t-score of 2.13281

```
Lower = x - 2.13281*s
Upper = x + 2.13281*s

Lower =   29.4594

Upper =   44.5406
```

With a sample size of 5, I predict the 90% confidence interval will be

$$\bar{x} - 2.13281s < roll < \bar{x} + 2.13281s$$

$$29.4594 < roll < 44.5406$$ p = 0.9, t-test

$$24.1591 < roll < 42.8409$$ normal approximation (problem #4)

This is a little off, but then it only uses a sample size of five

Problem 7: Using Matlab, determine ten values for Y

Y = 2d4 + 3d6 + 4d8

```
DATA = [];
for i=1:10
   d4 = ceil( 4*rand(2,1) );
   d6 = ceil( 6*rand(3,1) );
   d8 = ceil( 8*rand(4,1) );
   Y = sum(d4) + sum(d6) + sum(d8);
   DATA = [DATA, Y];
   end
DATA
x = mean(DATA)
s = std(DATA)
```

```
DATA =    31    35    38    32    42    33    39    34    32    24
x =    34
s =    4.9889
```

7a) From this, determine the mean and standard deviation of your data set.

see above

7b) Use a t-test to determine...

The probabillity of scoring more than 39.5 points

$$t = \left(\frac{39.5 - \bar{x}}{s}\right) = \left(\frac{39.5 - 34}{4.9889}\right) = 1.1025$$

From StatTrek, this corresponds to a probability of 14.943%

p = 14.943% t-test, sample size = 10

p = 14.537% normal pdf, sample size infinity

The 90% confidence interval

With 9 degrees of freedom, the t-score for 5% tails is

t = 1.83203

The 90% confidence interval is

$$\bar{x} - 1.83203s < roll < \bar{x} + 1.83203s$$

$$24.8602 < roll < 43.1398$$ p = 0.9, sample size = 10

$$24.1591 < roll < 42.8409$$ p = 0.9, sample size = infinity (problem 4)

Summary

The probability of Rolling 39.5 or more is

Method	p(y > 39.5)	# Rolls
Monte-Carlo	15.03%	100,000
Normal Approx	14.54%	0
t-Test	25.92%	5
t-Test	14.94%	10

The 90% confidene inteval is

Method	90% Confidence Interval	# Rolls
Monte-Carlo	[24, 43]	100,000
Normal Approx	(24.15, 42.54)	0
t-Test	(29.45, 44.54)	5
t-Test	(24.86, 43.13)	10

Using statistics, you can determine the same information without having to roll the dice 100,000 times

- If each experiment costs $10 to run, that can save a *lot* of money.