ECE 341 - Homework #14

Chi-Squared Tests.

In-Person vs. Online:

1) Use a Chi-Squared test to determine whether people who sign up for the in-person section of a class have a higher chance of geting an A.

	In-Person		Online	
	# A's	Total #	# A's	Total #
ECE 111	30	49	13	33
ECE 376	15	38	3	11
ECE 461	8	20	2	9

Treat the percentages of A's for in-person as the expected result

	p in-person A / total in-person	np	N	chi-squared
111 Online	30/49	20.2	13	2.5663
376 Online	15/38	4.34	3	0.4137
461 Online	8/20	3.6	2	0.7111
			Total	**3.6912**

From StatTrek, with 2 degrees of freedom, a chi-squared score of 3.6912 corresponds to a probability of 84.0%

There is an 84.0% chance that the chance to get an A is different for in-person and online students.

Loaded Dice

2) The following Matlab code generates 60 random numbers from 1..6 (sixty 6-sided dice)

```
RESULT = zeros(1,6);
for i=1:60
    d6 = ceil(rand*6);
    RESULT(d6) = RESULT(d6) + 1;
end
```

Use a chi-squared test to determine if this is a fair die.

Running the code:

```
RESULT =     8    11    15     8    10     8
```

Throwing this into a chi-squared table

die roll	p	np	N	chi-squared
1	1/6	10	8	0.4
2	1/6	10	11	0.1
3	1/6	10	15	2.5
4	1/6	10	8	0.4
5	1/6	10	10	0
6	1/6	10	8	0.4
			Total	**3.8**

From StatTrek, with 5 degrees of freedom, a chi-squared score of 3.8 corresponds to a probabilty of 0.42

I'm 42% certain that this is not a fair die (no conclusion)

3) The following Matlab code generates 60 random die rolls with 5% loading (5% of the time you roll a 6)

```
RESULT = zeros(1,6);
for i=1:60
    if(rand < 0.05) d6 = 6;
    else d6 = ceil(rand*6);
    end
    RESULT(d6) = RESULT(d6) + 1;
end
RESULT
```

Use a chi-squared test to determine if this is a fair die.

RESULT =   8   15   8   7   10   12

Setting up a chi-squared table

die roll	p	np	N	chi-squared
1	1/6	10	8	0.4
2	1/6	10	15	2.5
3	1/6	10	8	0.4
4	1/6	10	7	0.9
5	1/6	10	10	0
6	1/6	10	12	0.4
			Total	4.6

From StatTrek, a chi-squared score of 4.6 with 5 degrees of freedom corresponds to a probability of 53%

I'm 53% certain that this is not a fair die (no conclusion)

4) Repeat problem #3 with 600 die rolls.

```
RESULT = zeros(1,6);
for i=1:600
    if(rand < 0.05) d6 = 6;
    else d6 = ceil(rand*6);
    end
    RESULT(d6) = RESULT(d6) + 1;
end
RESULT

RESULT =    117    89    85    101    87    121
```

die roll	p	np	N	chi-squared
1	1/6	100	117	2.89
2	1/6	100	89	1.21
3	1/6	100	85	2.25
4	1/6	100	101	0.01
5	1/6	100	87	1.69
6	1/6	100	121	4.41
			Total	12.46

From StatTrek, I'm 97% certain that this is not a fair die

with enough data, you can detect small differences

Am I psychic?

5) Shuffle a deck of 52 playing cards. Without looking at the top card, predict the suit (clubs, diamonds, hearts, and spades). Repeat for all 52 cards, keeping track of how many you got right and how many you got wrong.

- From the results, use a chi-squared test to determine if you are just guessing (25% chance of getting the suit correct.)

Going through the deck, I guessed correctly 9 times out of 52 attempts

result	p	np	N	chi-squared
correct	1/4	13	9	1.2308
incorrect	3/4	39	43	0.4103
			Total	1.641

From StatTrek, with one degree of freedom, a chi-squared score of 1.641 corresponds to a probability of 80%

I'm 80% certain that I'm not just guessing.

my results are worst than just guessing (a.k.a. the monkey score), meaning I'm 80% certain that I'm unlucky

Central Limit Theorem:

6) The following code sums four uniform distributions

    $Y = \text{sum(rand(4,1))};$

The Central Limit Theorem states that this will converge to a normal distribution with

- mean = 2
- variance = 4/12

Use a chi-squared test to determine if Y does / does not have a normal distribution.

Split the data into 8 bins. Count how many times the data falls into each bin

```
Result = zeros(8,1);
for i=1:200
   y = sum(rand(4,1));
   n = ceil(y*2);
   Result(n) = Result(n) + 1;
end

Result
```

bin	y	cdf (StatTrek)	range	pdf	np	N	chi-squared
1	0.5	0.005	< 0.5	0.005	1	1	0
2	1	0.042	(0.5, 1.0)	0.037	7.4	8	0.0486
3	1.5	0.193	(1.0, 1.5)	0.151	30.2	34	0.4781
4	2	0.5	(1.5, 2.0)	0.307	61.4	57	0.3153
5	2.5	0.807	(2.0, 2.5)	0.307	61.4	65	0.2111
6	3	0.958	(2.5, 3.0)	0.151	30.2	29	0.0477
7	3.5	0.995	(3.0, 3.5)	0.037	7.4	6	0.2649
8	4	1	> 3.5	0.005	1	0	1
						Total	**2.3657**

From StatTrek, this corresponds to a probability of 0.06

I'm 6% certain that this is not a normal distribution (no conclusion)

with 200 data points, this looks pretty much like a normal distribution

Repeat with 2000 data points

bin	y	cdf (StatTrek)	range	pdf	np	N	chi-squared
1	0.5	0.005	< 0.5	0.005	10	5	2.5
2	1	0.042	(0.5, 1.0)	0.037	74	91	3.9054
3	1.5	0.193	(1.0, 1.5)	0.151	302	320	1.0728
4	2	0.5	(1.5, 2.0)	0.307	614	571	3.0114
5	2.5	0.807	(2.0, 2.5)	0.307	614	612	0.0065
6	3	0.958	(2.5, 3.0)	0.151	302	317	0.745
7	3.5	0.995	(3.0, 3.5)	0.037	74	79	0.3378
8	4	1	> 3.5	0.005	10	5	2.5
						Total	14.079

Now the probability of rejecting the null hypothesis (it's a normal distribution) is 95%

Repeat with 20,000 data points

bin	y	cdf (StatTrek)	range	pdf	np	N	chi-squared
1	0.5	0.005	< 0.5	0.005	100	45	30.25
2	1	0.042	(0.5, 1.0)	0.037	740	764	0.7784
3	1.5	0.193	(1.0, 1.5)	0.151	3,020	3,220	13.245
4	2	0.5	(1.5, 2.0)	0.307	6,140	6,073	0.7311
5	2.5	0.807	(2.0, 2.5)	0.307	6,140	5,957	5.4542
6	3	0.958	(2.5, 3.0)	0.151	3,020	3,082	1.2728
7	3.5	0.995	(3.0, 3.5)	0.037	740	803	5.3635
8	4	1	> 3.5	0.005	100	56	19.36
						Total	76.4551

Now the probability of rejecting the null hypothesis (it's a normal distribution) is > 99.995%

With enough data, you *can* spot the difference

Poisson approximation for a binomial distribution.

7) Let X be the number of 1's you get when you roll 60 dice. The Poisson approximation for the pdf is

$$\binom{60}{x}\left(\tfrac{1}{6}\right)^{x}\left(\tfrac{5}{6}\right)^{60-x} \approx \left(\tfrac{1}{x!}\right)10^{x}\,e^{-10}$$

- Use Matlab to count the number of 1's you get when you roll 60 dice
- Repeat 200 times
- Check whether the result is consistent with a Poisson distribution with $\lambda = Np = 10$ using a Chi-squred test
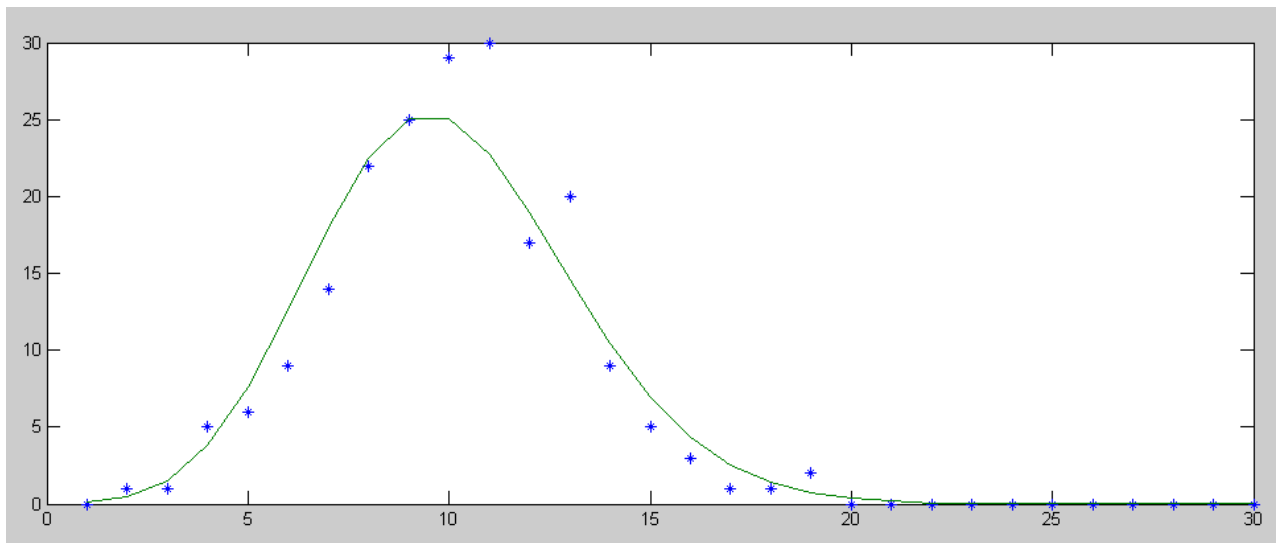
Matlab Code:

```
Result = zeros(60,1);
for i=1:200
    Dice = ceil(6*rand(60,1));
    N = sum(Dice == 1);
    Result(N) = Result(N) + 1;
end

k = [1:30]';
Result = Result(k);

Poisson = 1./ factorial(k) .* (10.^k) * exp(-10) * 200;

plot(k,Result,'*',k,Poisson)

Chi2 = ( (Poisson - Result).^2 ./ Poisson );
sum(Chi2)
```


The chi-squared score is 13.75

From StatTrek, a chi-squared score of 13.75 with 29 degrees of freedom corresponds to a probability of 0.007

With only 200 data points, it's hard to tell the difference between a Poisson distribution and a binomial distribution