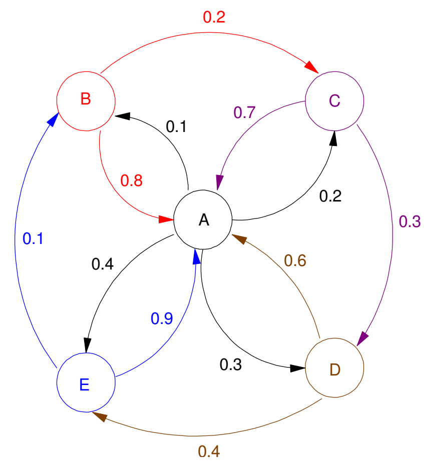
ECE 341 - Test #3: Name ____________

Markov Chains and Data Analysis

1) Markov Chains: Assume five players are tossing a ball around.

- Each second the player with the ball tosses it.
- The probability that the player tosss the ball to someone else is shown below.
- At k=0, player A has the ball.



a) Express the probability that a player has the ball after k tosses as:

$$X(k+1) = MX(k)$$

where X(k) is the probability that player {A, B, C, D, E} has the ball at toss #k.

$$\begin{bmatrix} A(k+1) \\ B(k+1) \\ C(k+1) \\ D(k+1) \\ E(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 0.8 & 0.7 & 0.6 & 0.9 \\ 0.1 & 0 & 0 & 0 & 0.1 \\ 0.2 & 0.2 & 0 & 0 & 0 \\ 0.3 & 0 & 0.3 & 0 & 0 \\ 0.4 & 0 & 0 & 0.4 & 0 \end{bmatrix} \begin{bmatrix} A(k) \\ B(k) \\ C(k) \\ D(k) \\ E(k) \end{bmatrix}$$

Note:

- Columns add to one (the ball goes somewhere)

b) Determine the probability that A has the ball after 10 tosses.

$$X(10) = M^{10} X(0)$$

Assume A starts with the ball

$$X(10) = \begin{bmatrix} 0 & 0.8 & 0.7 & 0.6 & 0.9 \\ 0.1 & 0 & 0 & 0 & 0.1 \\ 0.2 & 0.2 & 0 & 0 & 0 \\ 0.3 & 0 & 0.3 & 0 & 0 \\ 0.4 & 0 & 0 & 0.4 & 0 \end{bmatrix}^{10} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In Matlab

```
>> M = [0,0.8,0.7,0.6,0.9 ; 0.1,0,0,0,0.1 ; 0.2,0.2,0,0,0 ; 0.3,0,0.3,0,0 ;
0.4,0,0,0.4,0]

         0    0.8000    0.7000    0.6000    0.9000
    0.1000         0         0         0    0.1000
    0.2000    0.2000         0         0         0
    0.3000         0    0.3000         0         0
    0.4000         0         0    0.4000         0

>> X0 = [1,0,0,0,0]'

     1
     0
     0
     0
     0

>> M^10 * X0

a    0.4676
b    0.0644
c    0.0922
d    0.1503
e    0.2255
```

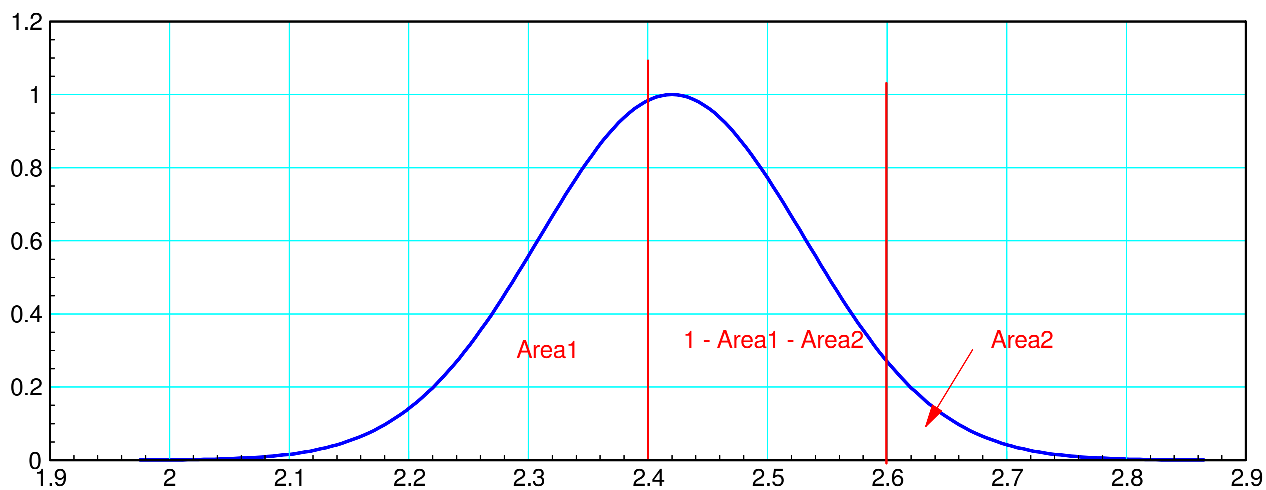
There is a 46.76% chance A has the ball after 10 tosses

2) t-Test (One data set). A 4-sided die may or may not be loaded. If it's a fair die, the mean of the die rolls should be 2.5.

- Use a t-test to determine the probability that the mean of the die is in the range of (2.4, 2.6)
- (i.e. is this a fair die?)
- note: This is population question

1's	2's	3's	4's	mean	st dev	# rolls
29	20	31	20	2.4200	1.1117	100
$p(2.4 < \mu < 2.6) = 51.7\%$						

Solution: Find the t-score and the probability at 2.4 and 2.6. It sometimes helps to draw the pdf:



To find the area between (2.4, 2.6),

- Find the area of the tail, left of 2.4
- Find the area of the tail, right of 2.6
- The remaining area is the probability that the mean is in the range of (2.4, 2.6)

Since this is a population question, divide the variance by n

Area left of 2.4:

$$t = \left(\frac{2.4 - \bar{x}}{s/\sqrt{n}}\right)$$

$$t = \left(\frac{2.4 - 2.42}{\frac{1.1117}{\sqrt{100}}}\right) = -0.1799$$

From a t-table with 99 degrees of freedom, the area to the left is 0.429

Area right of 2.6:

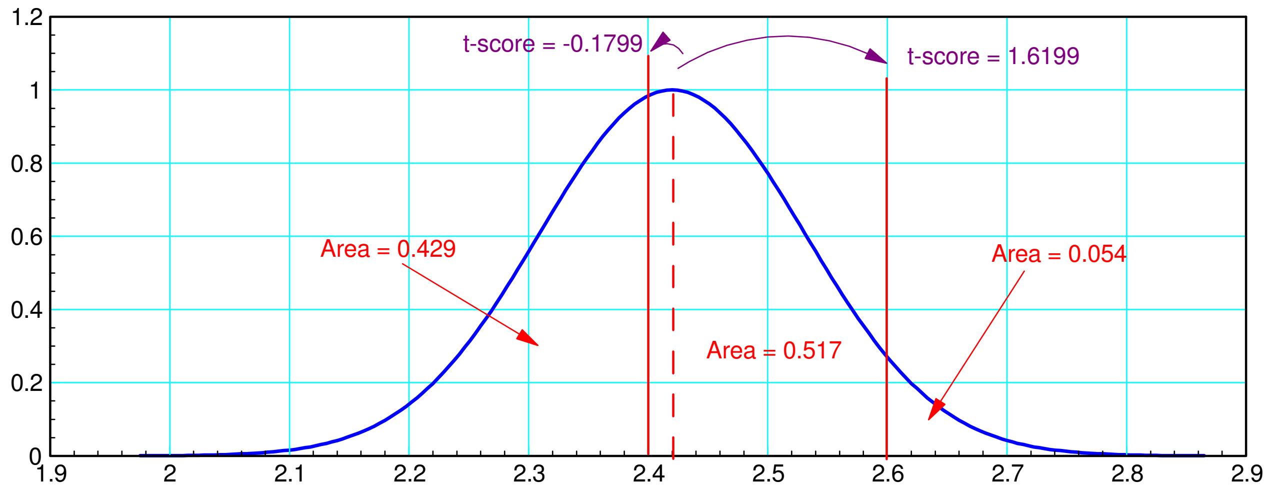
$$t = \left(\frac{2.6 - 2.42}{\frac{1.1117}{\sqrt{100}}} \right) = 1.6191$$

From a t-table with 99 degrees of freedom, the area of the tail is 0.054

The area in the middle is then

p = 1 - 0.054 - 0.429 = 0.517

There is a 51.7% chance the population's mean in in the range of (2.4, 2.6)



3) t-Test (Two data sets): Two four-sided dice are rolled N times. They might be fair dice, they might both be loaded dice.

Determine the probability that the mean of die A is within 0.1 of the mean of B

- i.e. $-0.1 < \mu_a - \mu_b < 0.1$
- note: this is a population question

	1's	2's	3's	4's	mean	st dev	# rolls
A	29	20	31	20	2.4200	1.1117	100
B	14	23	22	21	2.6250	1.0599	80
$p(-0.1 < \mu_a - \mu_b < +0.1) = 22.8\%$							

Define a new variable, W, which is the difference between A and B. Since this is a population question, divide the variance by the sample size:

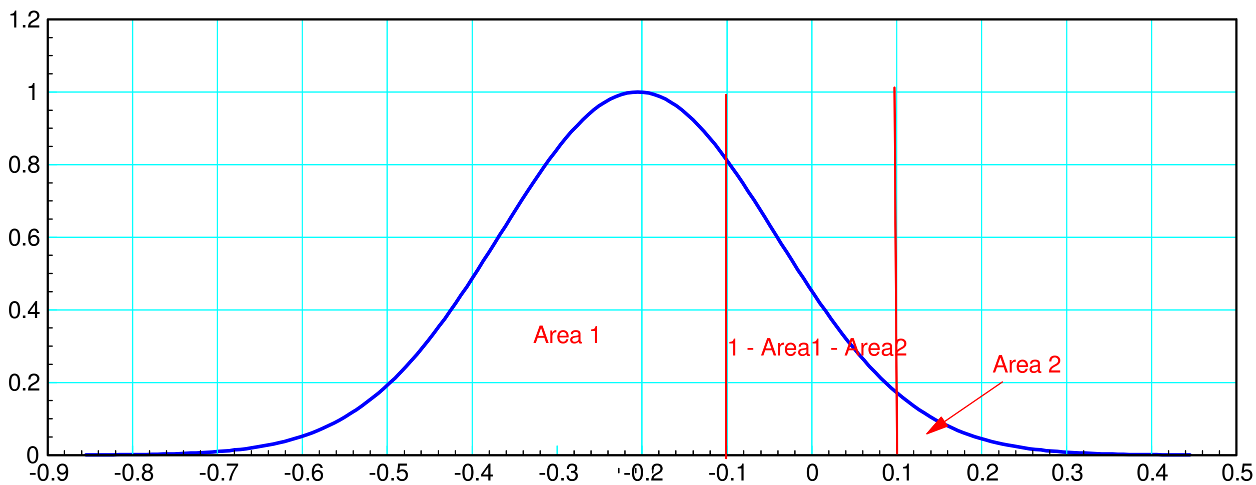
W = A - B

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = -0.205$$

$$s_w = \sqrt{\frac{s_a^2}{100} + \frac{s_b^2}{80}} = 0.162484$$

It helps to draw the pdf. The are between (-0.1, +0.1) is one minus

- The area of the tail left of -0.1, minus
- The area of the tail right of +0.1



Find the t-score at -0.1

$$t = \left(\frac{-0.1-(-0.205)}{0.162484} \right) = 0.646217$$

The area to the left is 0.7400 (80 degrees of freedom)

Find the t-score at +0.1

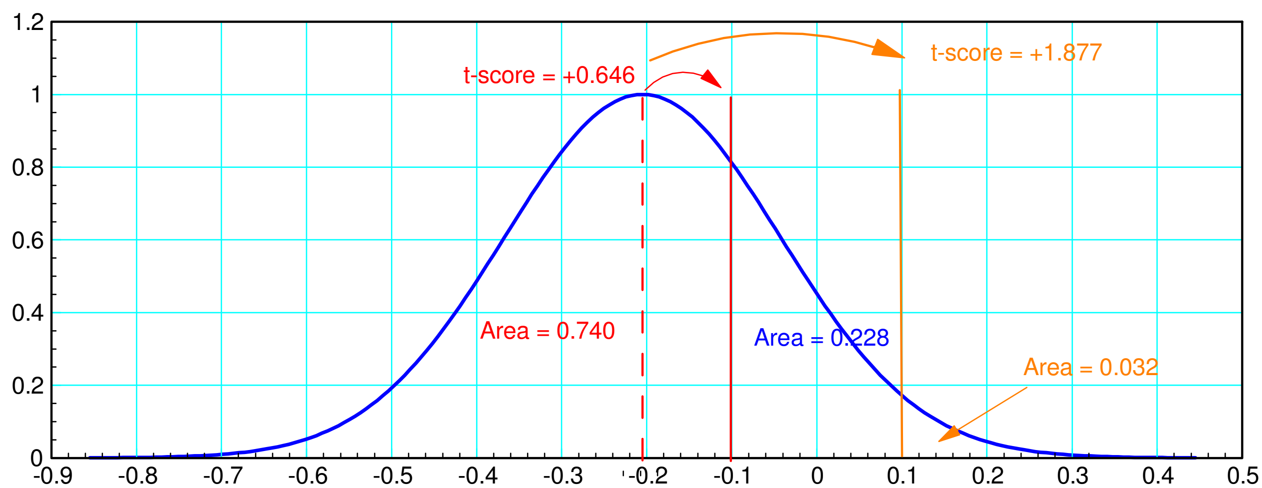
$$t = \left(\frac{+0.1-(-0.205)}{0.162484} \right) = 1.877105$$

The area to the right is 0.032

The area in the middle is the remainder:

p = 1.000 - 0.032 - 0.740 = 0.228

There is a 22.8% chance the mean of A and B are within 0.1 of eachother



4) Chi-Squared Test: A 4-sided die is rolled 100 times. Determine if this is a fair die using a chi-squared test

1's	2's	3's	4's	mean	st dev	# rolls
29	20	31	20	2.4200	1.1117	100

die roll	p	np	N	chi-squared
1	0.25	25	29	0.64
2	0.25	25	20	1
3	2.5	25	31	1.44
4	0.25	25	20	1
			Total	**4.08**

From StatTrek, with three degrees of freedom, this corresponds to a probability of 0.74705

There is a 74.705% chance the die is not fair

5) ANOVA (Three data sets): Three 4-sided dice are rolled. They may or may not be loaded.

Use an ANOVA test to determine if the three dice have the same mean

	1's	2's	3's	4's	mean	st dev	# rolls
A	29	20	31	20	2.4200	1.1117	100
B	14	23	22	21	2.6250	1.0599	80
C	20	12	14	14	2.3667	1.1784	60

In Matlab

```
Xa = 2.420;
Sa = 1.1117;
Xb = 2.6250;
Sb = 1.0599;
Xc = 2.3667;
Sc = 1.1784;
Na = 100;
Nb = 80;
Nc = 60;
k = 3;
N = Na + Nb + Nc
G = (Na*Xa + Nb*Xb + Nc*Xc) / N
MSSb = (Na*(Xa-G)^2 + Nb*(Xb-G)^2 + Nc*(Xc-G)^2) / (k-1)
MSSw = ((Na-1)*Sa^2 + (Nb-1)*Sb^2 + (Nc-1)*Sc^2) / (N-k)
F = MSSb / MSSw

N    =  240
G    =    2.4750
MSSb =    1.4031
MSSw =    1.2364
F    =    1.1348
```

From StatTrek, this corresponds to a probability of 66.7%

- $F = 1.1348$
- numberator has 2 (k-1) degrees of freedom
- denominator has 237 (N-k) degrees of freedom
- $p = 0.677$

There is a 66.7% chance that the populations have different means (one or more die is loaded)