# ECE 341 - Test #3: Name ____________

Markov Chains and Data Analysis

**1) Markov Chains:** Two players are plaing a match. To win the match, a player has to win three games in a row. Assume for each game

- A has a 50% chance of winning
- B has a 40% chance of winning, and
- There is a 10% chance of a tie (breaking all winning streaks)

a) Express the probability that player A or B has an n-game winning streak as:

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

where X(k) is defined as the probability after game k that

$$\begin{bmatrix} P3(k+1) \\ P2(k+1) \\ P1(k+1) \\ P0(k+1) \\ M1(k+1) \\ M2(k+1) \\ M3(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 \\ 0 & 0.4 & 0.4 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 1 \end{bmatrix} \begin{bmatrix} P3(k) \\ P2(k) \\ P1(k) \\ P0(k) \\ M1(k) \\ M2(k) \\ M3(k) \end{bmatrix}$$

b) Explain how you would determine the odds of A winning the match (A is first to get a 3-game winning streak).

- Or find the odds if you have access to Matlab

In Matlab

```
>> a1 = [1,0.5,0,0,0,0,0];
>> a2 = [0,0,0.5,0,0,0,0];
>> a3 = [0,0,0,0.5,0.5,0.5,0];
>> a4 = [0,0.1,0.1,0.1,0.1,0.1,0];
>> a5 = [0,0.4,0.4,0.4,0,0,0];
>> a6 = [0,0,0,0,0.4,0,0];
>> a7 = [0,0,0,0,0,0.4,1];
>> A = [a1;a2;a3;a4;a5;a6;a7]

    1.0000    0.5000         0         0         0         0         0
         0         0    0.5000         0         0         0         0
         0         0         0    0.5000    0.5000    0.5000         0
         0    0.1000    0.1000    0.1000    0.1000    0.1000         0
         0    0.4000    0.4000    0.4000         0         0         0
         0         0         0         0    0.4000         0         0
         0         0         0         0         0    0.4000    1.0000

>> X0 = [0;0;0;1;0;0;0]

X0 =

     0
     0
     0
     1
     0
     0
     0

>> A^100 * X0

ans =

    0.6352
    0.0000
    0.0000
    0.0000
    0.0000
    0.0000
    0.3648
```

**A has a 63.52% chance of winning the match**

**2) t-Test (One data set).** The gain of nine 6144 NPN transistors were recorded:

- Gain = {374, 370, 359, 370, 351, 357, 352, 364, 372}
- mean = 363.2222
- standard deviation = 8.7860

a) (individual) If I measure a tenth 6144 NPN transistor, what is the 90% confidence interval for the gain of this transistor?

5% tails with 8 degrees of freedom corresponds to a t-score of -1.85938

The 90% confidence interval is

$$\bar{x} - 1.85938s < gain < \bar{x} + 1.85938s$$

346.8855 <Individual Gain< 379.5585

```
>> x = 363.222;
>> s = 8.7860;
>> high = x + 1.85938*s

high =  379.5585

>> low = x - 1.85938*s

low =  346.8855
```

b) (population) What is the 90% confidence interval for the average gain of a 6144 NPN transistor?

For populations, you divide the variance by the sample size

$$\bar{x} - 1.66406\left(\frac{s}{\sqrt{9}}\right) < gain < \bar{x} + 1.66406\left(\frac{s}{\sqrt{9}}\right)$$

357.7765 <Population Gain< 368.6675

```
>> high = x + 1.85938*s/sqrt(9)

high =  368.6675

>> low = x - 1.85938*s/sqrt(9)

low =  357.7765

>>
```

**3) t-Test (Two data sets):** The points scored by the Vikings and Packers in 2022 are as follows:

| Team | Average points | standard deviation points | games |
| --- | --- | --- | --- |
| Vikings | 24.9444 | 8.3135 | 17 |
| Packers | 21.7647 | 9.2096 | 17 |

a) (Individual) Determine using a Student t-Test the probability that the Vikings will outscore the Packers next time they play.

Form a new variable, W = V - P

$$\bar{x}_w = \bar{x}_v - \bar{x}_p = +3.1797$$

$$s_w = \sqrt{s_v^2 + s_p^2} = 12.4096$$

Find the t-score

$$t = \frac{x_w}{s_w} = 0.2563$$

Convert to a probability with 16 degrees of freedom

p = 0.59951

**The Vikings have a 59.051% chance of scoring more points than the Packers in their next game**

```
>> Xv = 24.9444;
>> Sv = 8.3135;
>> Xp = 21.7647;
>> Sp = 9.2096;
>> Xw = Xv - Xp

Xw =    3.1797

>> Sw = sqrt(Sv^2 + Sp^2)

Sw =   12.4069

>> t = Xw / Sw

t =    0.2563

>>
```

b) (population) Determine using a Student t-Test the probability that the Vikings have the better offense overall.

For populations, you divide the variace by the sample size

```
>> Xw = Xv - Xp

Xw =    3.1797

>> Sw = sqrt( (Sv^2)/17 + (Sp^2)/17)

Sw =    3.0091

>> t = Xw / Sw

t =    1.0567
```

From StatTrek, a t-score of 1.0567 with 16 degrees of freedom corresponds to a probability of 084683

**The Vikings have the better offense with a probability of 85.693%**

4) Chi-Squared Test:  The points scored by the Minnesota Vikings in 2022 were:

| Point Range | 0-9 | 10-19 | 20-29 | 30-39 | 40-49 | Total Games |
|---|---|---|---|---|---|---|
| Frequency | 3 | 2 | 12 | 4 | 0 | 21 |

Use a chi-squared test to determine the chance that the Vikings score has a uniform distribution (equal liklihood over the range of (0,49) points).

Set up a chi-squared table

| bin | p | np | N | chi-squared |
|---|---|---|---|---|
| 0-9 | 1/5 | 4.2 | 3 | 0.34 |
| 10-19 | 1/5 | 4.2 | 2 | 1.15 |
| 20-29 | 1/5 | 4.2 | 12 | 14.49 |
| 30-39 | 1/5 | 4.2 | 4 | 0.01 |
| 40-49 | 1/5 | 4.2 | 0 | 4.2 |
| | | | **Total** | **20.19** |

Convert to a probability using a chi-squared table with four degrees of freedom

$p = 0.99954$

**There is a 99.954% chance that the Viking's score does not follow a uniform distribution**

5) ANOVA (Three data sets): The points scored in 20222 by the Vikings, Packers, and Bears are as follows. Use an Analysis of Variance test to determine the probability that the means are different.

| Team | mean (points per game) | standard deviation (points per game) | # games |
|---|---|---|---|
| A: Vikings | 24.9444 | 8.3135 | 17 |
| B: Packers | 21.7647 | 9.2096 | 17 |
| C: Bears | 19.1765 | 8.4575 | 17 |

```
Xa = 24.9444;
Sa = 8.3135;
Xb = 21.7647;
Sb = 9.2096;
Xc = 19.1765;
Sc = 8.4575;
Na = 17;
Nb = 17;
Nc = 17;
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 =       51
G =       21.9619
MSSb =  141.8875
MSSw =   75.1534
F =       1.8880

From an F-table with 2 degrees of freedom (numerator) and 48 degrees of freedom
(denominator), this corresponds to a probability of 0.83756
```

**There is an 83.756% chance that the means of these populations are different**