ECE 341 - Homework #10

Testing with Normal Distributions. Summer 2023

Testing with Normal Distributions

Assume the monthly temperatures in Fargo, ND are normal distributions with the following mean and standard deviation:

Monthly Low (Degrees F: Fargo, ND)												
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
mean	-23.8725	-20.6238	-8.1475	15.1775	27.3413	40.425	46.4875	43.3387	30.6763	19.15	-1.0875	-17.025
st dev	8.2179	7.8559	10.0237	7.0423	4.3864	4.1576	4.0938	4.1522	4.8861	5.5212	9.0417	9.1069

1) How cold will this November get

- With a confidence level of 80%?
- With a confidence level of 99%?
- With a confidence level of 100%?

For the low in November, the mean and standard deviation are

$$\mu = -1.0875F$$
$$\sigma = 9.0471F$$

80% confidence interval:

- The z-score for 10% tails is 1.28155

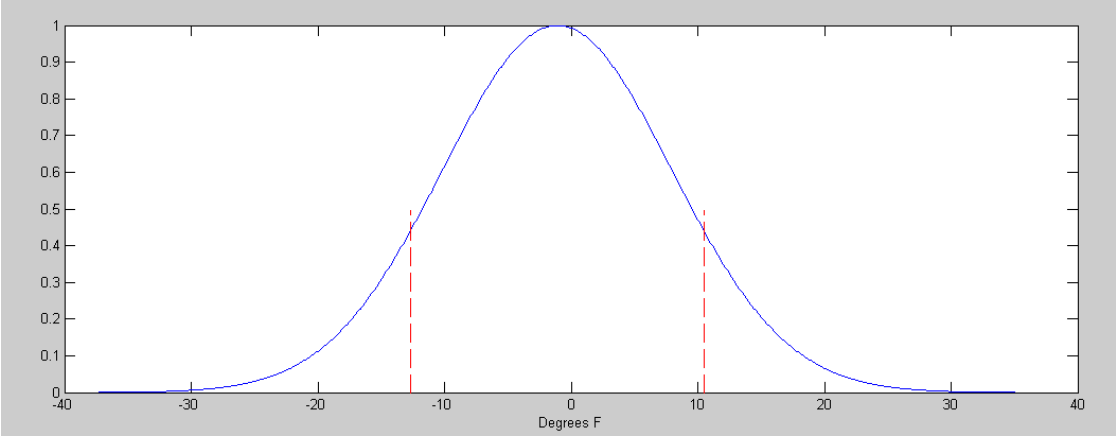
$$\mu - 1.28155\sigma < Low < \mu + 1.28155\sigma$$

$$-12.7175F < Low < 10.5425F$$

The low in November will be in the range of (-12.7175F, +10.5425F) with a probability of 80%

Just for fun, plot this

```
>> s1 = [-4:0.01:4]';
>> p = exp(-s1.^2 / 2);
>> x = -1.0875;
>> s = 9.0471;
>> plot(s1*s+x,p,'b',-12.7125*[1,1],[0,0.5],'r--',10.5425*[1,1],[0,0.5],'r--')
>> xlabel('Degrees F');
```



99% confidence interval:

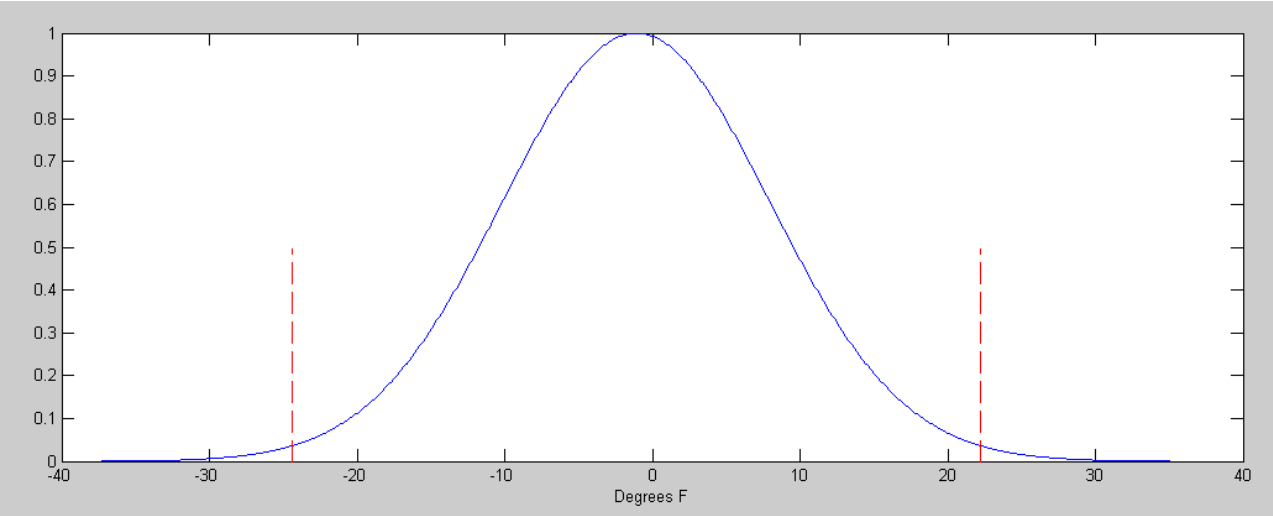
- The z-score for 0.5% tails is 2.5783

$$\mu - 2.5783\sigma < Low < \mu + 2.5783\sigma$$

$$-24.4136F < Low < +22.2386F$$

The low in November will be in the range of (-24.4136F, +22.2386F) with a probability of 99%

```
>> plot(s1*s+x,p,'b',-24.41*[1,1],[0,0.5],'r--',-22.23*[1,1],[0,0.5],'r--')
>> xlabel('Degrees F');
```



99% Confidence Interval

100% confidence interval: The z-score is infinity

The low in November will be in the range of (-infinity, +infinity) with a probability of 100%

- note: 100% probability is nonsense.
- Nothing is 100% certain

Monthly Low (Degrees F: Fargo, ND)												
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
mean	-23.8725	-20.6238	-8.1475	15.1775	27.3413	40.425	46.4875	43.3387	30.6763	19.15	-1.0875	-17.025
st dev	8.2179	7.8559	10.0237	7.0423	4.3864	4.1576	4.0938	4.1522	4.8861	5.5212	9.0417	9.1069

2) What is the probability that it will break -40F this coming January?

For January, the low has a mean and standard deviation of

$$\mu = -23.8725F$$
$$\sigma = 8.2179F$$

The z-score for -40F is

$$z = \left(\frac{-40F - \mu}{\sigma}\right) = -1.9625$$

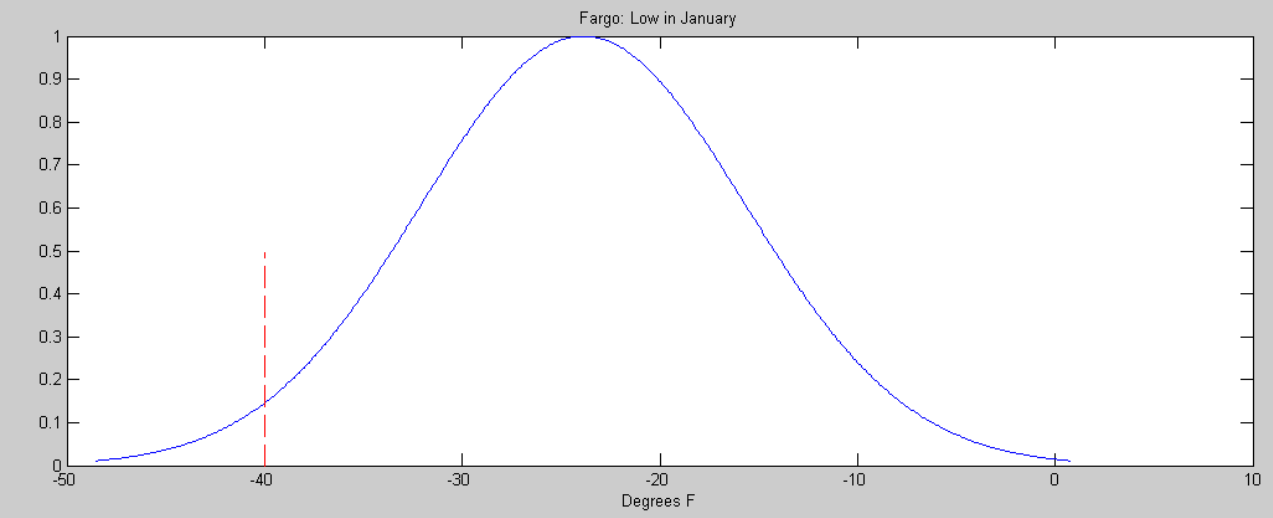
From StatTrek (or a normal table), the area of the tail with this z-score is

p = 0.02485

There is a 2.485% chance it will get colder than -40F this coming January

Plotting (just for fun)

```
>> s = [-3:0.01:3]';
>> p = exp(-s.^2 / 2);
>> plot(s*8.2179-23.8725, p, 'b', -40*[1,1],[0,0.5],'r--');
>> xlabel('Degrees F');
>> title('Fargo: Low in January');
```



Monthly High (Degrees F: Fargo, ND)												
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
mean	38.5363	41.0038	56.0625	78.1	87.8625	92.0138	94.625	94.6262	89.575	79.5	59.425	41.7875
st dev	6.4057	7.1528	10.7118	7.7909	4.5472	4.5281	4.0043	4.5967	5.6294	6.7842	7.4728	6.5327

3) What is the probability that it will break +100F in June?

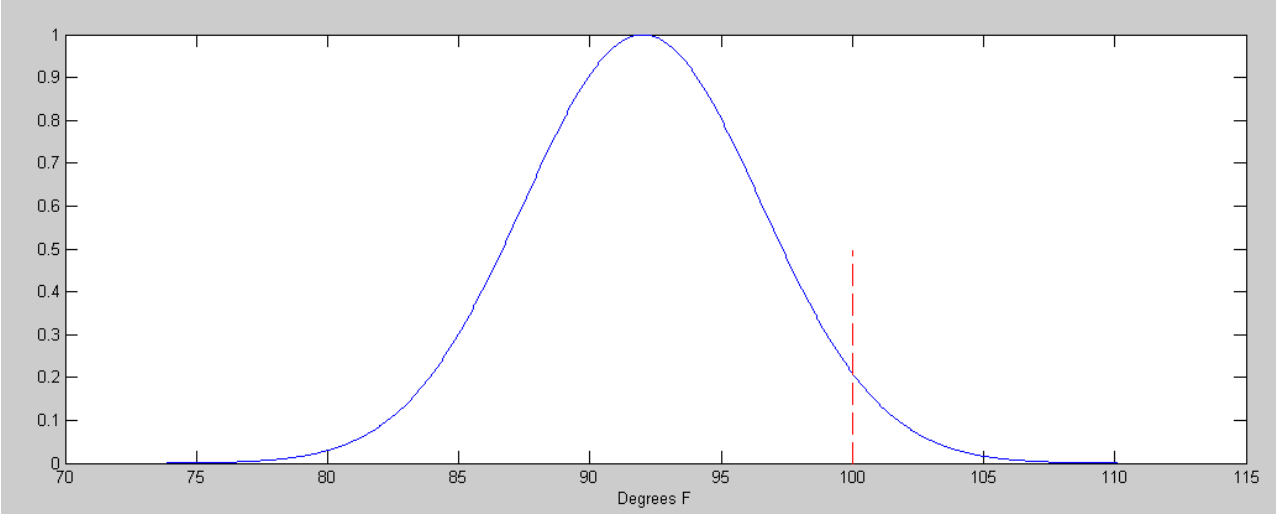
Calculate the z-score

$$z = \left(\frac{+100F - \mu}{\sigma}\right) = \left(\frac{100F - 92.0138F}{4.5281F}\right) = 1.7637$$

The area of the tail for this z-score (from a normal table or StatTrek) is 0.03889

There is a 3.889% chance of of breaking 100F in June

```
>> x = 92.01;
>> s = 4.5281;
>> plot(s1*s+x,p,'b',100*[1,1],[0,0.5],'r--')
>> xlabel('Degrees F');
```



High in June: 3.889% chance of breaking 100F

Monthly High (Degrees F: Fargo, ND)												
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
mean	38.5363	41.0038	56.0625	78.1	87.8625	92.0138	94.625	94.6262	89.575	79.5	59.425	41.7875
st dev	6.4057	7.1528	10.7118	7.7909	4.5472	4.5281	4.0043	4.5967	5.6294	6.7842	7.4728	6.5327

Testing with Two Populations

4) What is the probability that June will be warmer than July in a given year?

Create a new distribution, W

W = June - July

The mean & standard deviation are

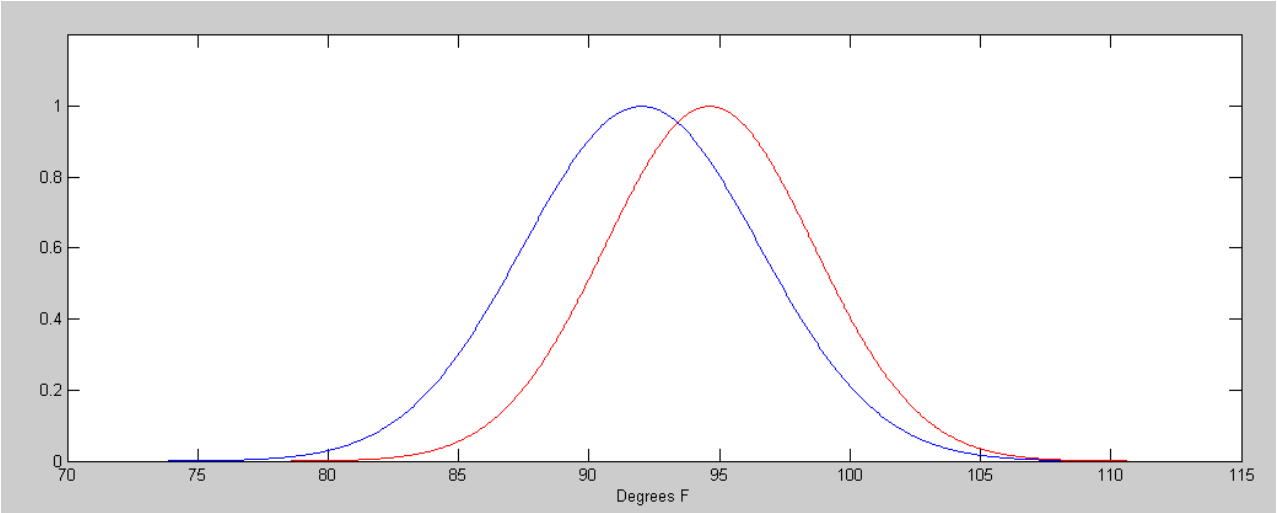
```
>> xa = 92.0138;
>> sa = 4.5281;
>> xb = 94.625;
>> sb = 4.0043;
>> Xw = xa - xb

Xw =   -2.6112

>> Sw = sqrt(sa^2 + sb^2)

Sw =    6.0447
```

The z-score for W > 0 is

$$z = \left(\frac{\mu_w - 0}{\sigma_w}\right) = -0.4320$$

From StatTrek, the tail for this z-score is 0.33287

There is a 33.287% chance the the high for June will be more than the high for July



pdf for June's high (blue) and July's high (red)

Monthly Low (Degrees F: Fargo, ND)												
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
mean	-23.8725	-20.6238	-8.1475	15.1775	27.3413	40.425	46.4875	43.3387	30.6763	19.15	-1.0875	-17.025
st dev	8.2179	7.8559	10.0237	7.0423	4.3864	4.1576	4.0938	4.1522	4.8861	5.5212	9.0417	9.1069

The low for 20 months are as follows:

```
{ -19.8, -19.0 -15.0 -11.0 -5.0 -5.0 -3.0 -2.0 0.0 3.0 4.0 8.0 9.0 11.0 14.0 15.0
15.0 16.0 21.2 23.0 }
```

5) Which months are March and which ones are April? What threshold do you use for separating the data?

Pick the midpoint as the separation between March and April

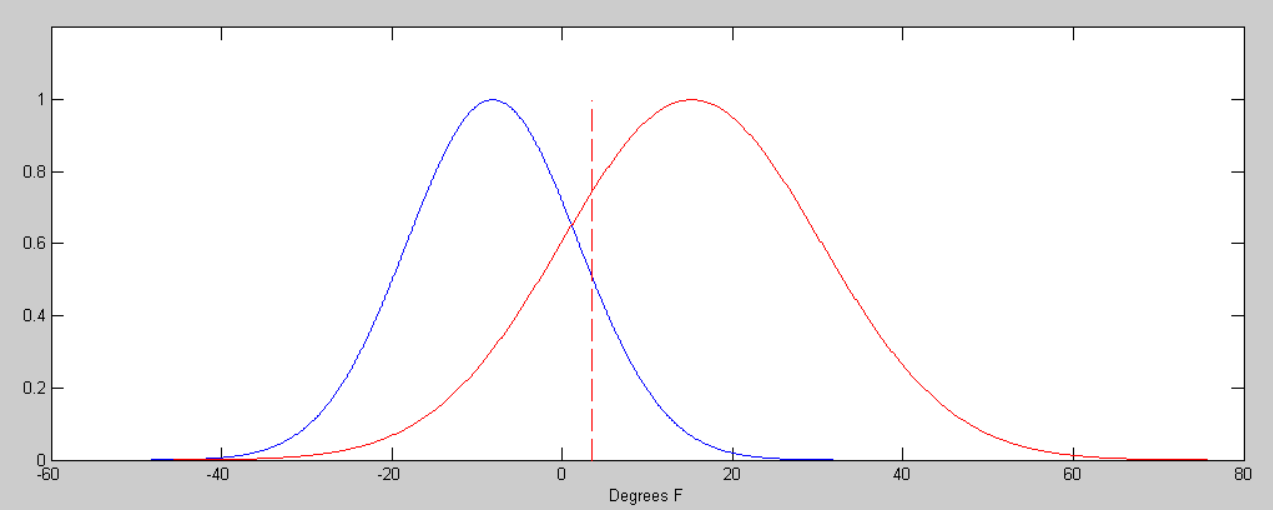
- If the low is less than +3.51F, call this month March
- If the low is more than +3.51F, call this month April

You could pick a different temperature as the threshold as well. Where you place this line determines the probabilities of false positives and negatives

March Months: (blue)

April Months (red)

```
{ -19.8, -19.0 -15.0 -11.0 -5.0 -5.0 -3.0 -2.0 0.0 3.0 4.0 8.0 9.0 11.0 14.0 15.0
15.0 16.0 21.2 23.0 }
```



pdf for the low in March (blue) and April (red) & threshold

6) With your threshold, what is the probability of

- A false positive ? (the temperature was assigned to March but actually came from April)
- A false negative? (the temperature was assigned to April but actually came from March)

False Positive: The area of the red curve to the left of the threshold

$$z = \left(\frac{3.51 - \mu_{April}}{\sigma_{April}} \right) = -1.6568$$

From StatTrek, this has an area of 0.04878

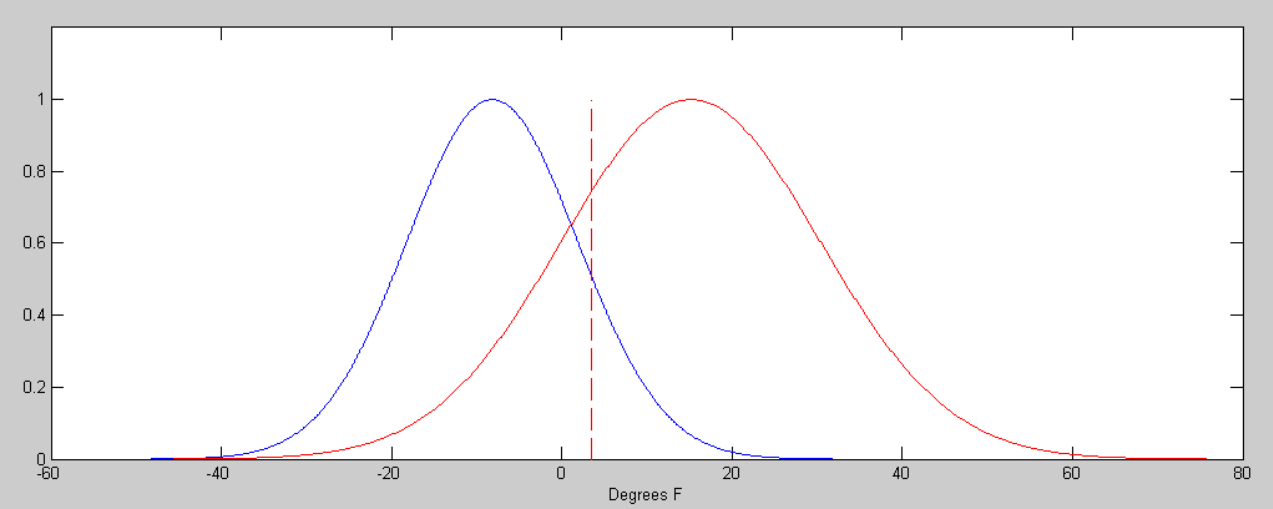
The probability of a false negative is 4.878%

False Negative: The are of the blue curve to the right of the threshold

$$z = \left(\frac{3.51 - \mu_{March}}{\sigma_{March}} \right) = 1.1630$$

From StatTrek, this has an area of 0.12241

The probability of a false positive is 12.241%



Regression Analysis

The average temperature in August in Fargo, ND is available at

`http://www.bisonacademy.com/ECE111/Code/Fargo_Weather_Monthly_Avg.txt`

7) Find the least-squares curve fit for this data as

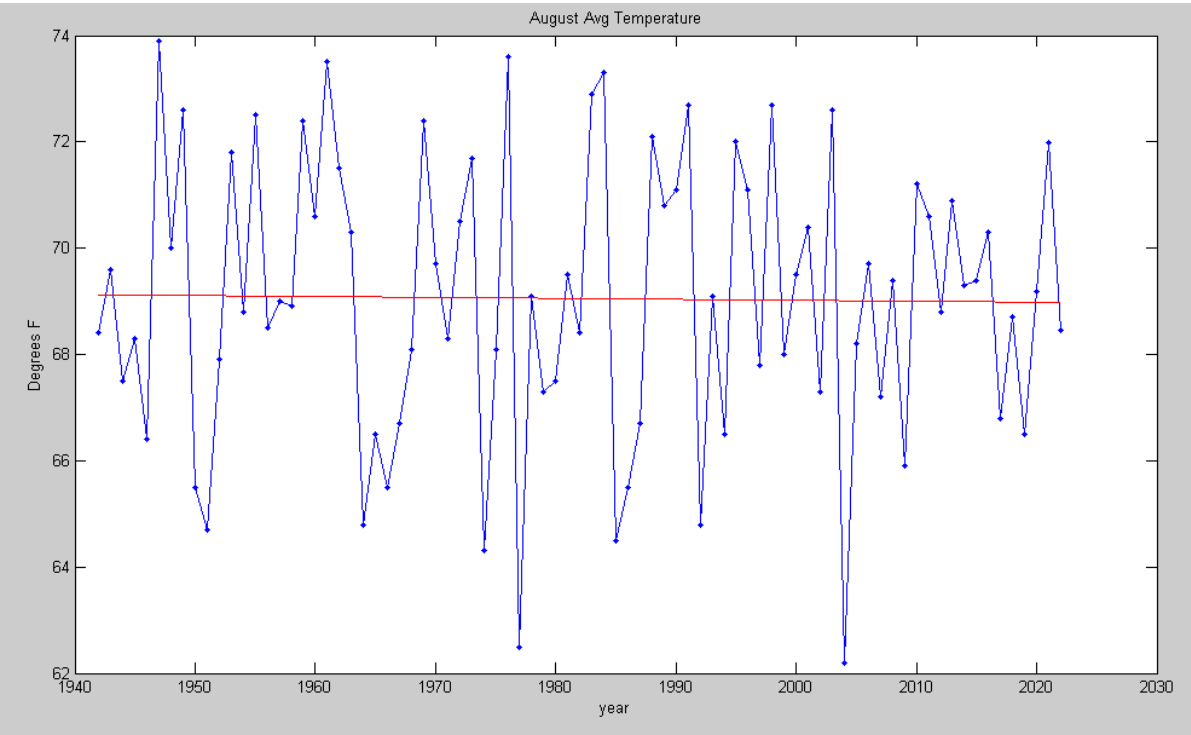
$$F = ay + b$$

where T is the temperature in degrees F and y is the year.

From this curve fit, how much has August in Fargo warmed up since 1942?

```
DATA = [ <paste data> ];
>> year = DATA(:,1);
>> Aug = DATA(:,9);
>> B = [year, year.^0];
>> A = inv(B'*B)*B'*Aug

a    -0.0018
b    72.6138

>> plot(year,Aug,'b.-',year,B*A,'r')
>> a = A(1);
>> a*80
     -0.1438
```

August has gotten slightly cooler over the past 80 years



8) Determine the correlation coefficient between

- The average temperature in May and June
  if May is hot, is June going to be hot?
- The average temprature in May and December
  if May is hot, is December going to be hot?

```
>> May = DATA(:,6);
>> June = DATA(:,7);
>> Dec = DATA(:,13);
>> Cov = mean(May .* June) - mean(May)*mean(June)

Cov =    2.7470

>> Correlation = Cov / ( std(May) * std(June) )

Correlation =    0.2381
```

There is a 23.81% correlation between the average temperature in May and June

- **It is slightly more likely to have a hot June if we had a hot May**

```
>> Cov = mean(May .* Dec) - mean(May)*mean(Dec)

Cov =   -1.3916

>> Correlation = Cov / ( std(May) * std(Dec) )

Correlation =   -0.0601
```

There is a negative 6.01% correlation between May and December

- **It is slightly more likely to have a cold December if we had a hot May**