Student t Distribution with 2 Populations

ECE 341: Random Processes

Lecture #24a

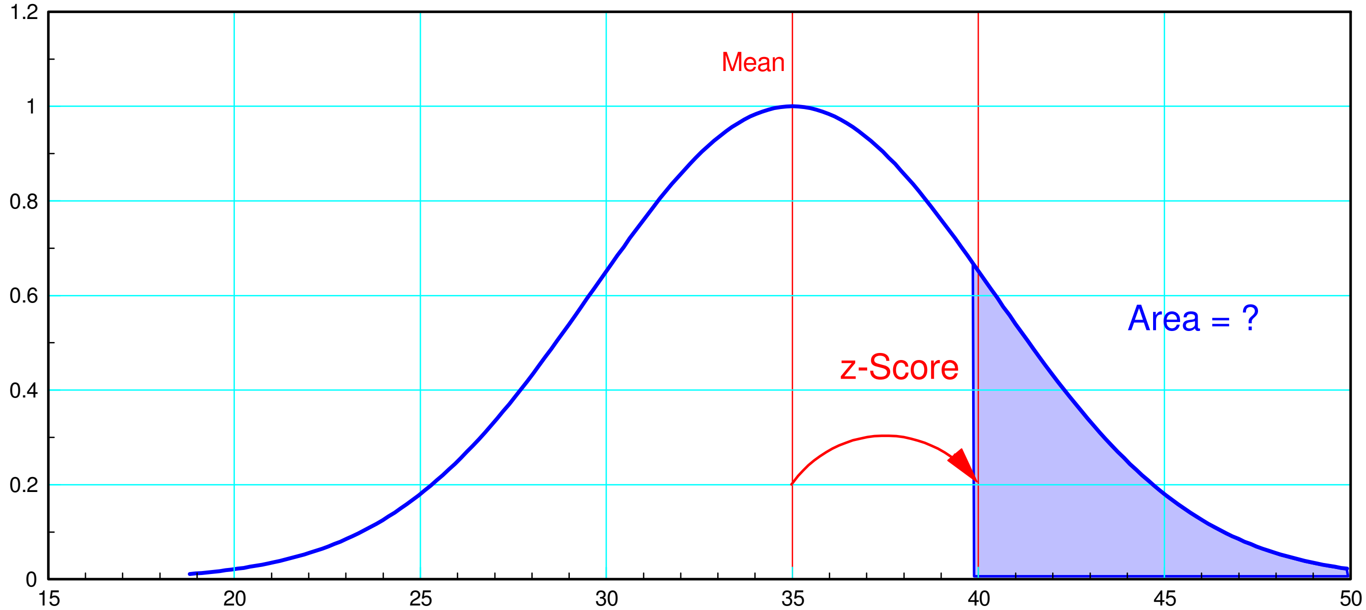
note: All lecture notes, homework sets, and solutions are posted on www.BisonAcademy.com

Recap: Testing with a Normal Approximation

- Assumes a single population
- Mean & standard deviation are known
- Example: What is the probability that the sum of 10d6 > 39.5?

Procedure

- Find the z-score
- Convert to a probability using a standard-normal table



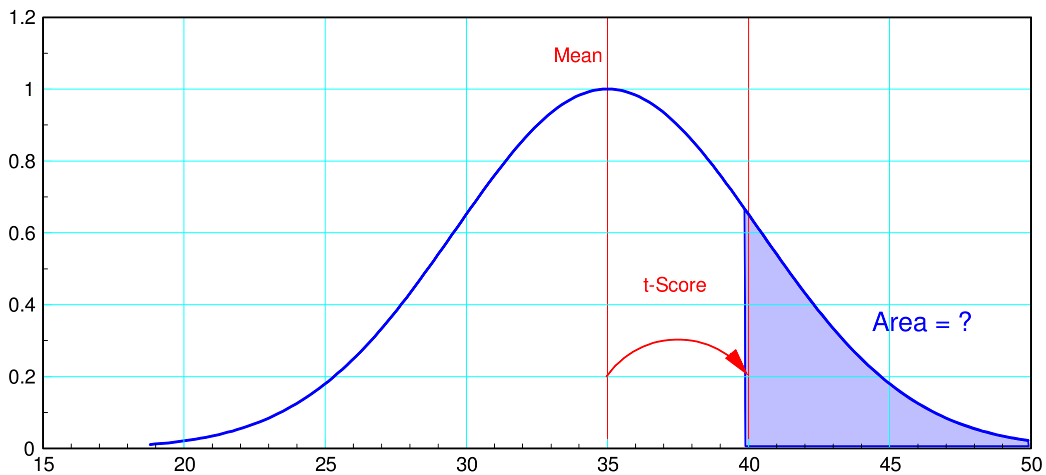
Recap: Testing with a Student-t Distribution

Assumes

- A single population
- A normal pdf
- N measurements with mean & standard deviation unknown

Examples:

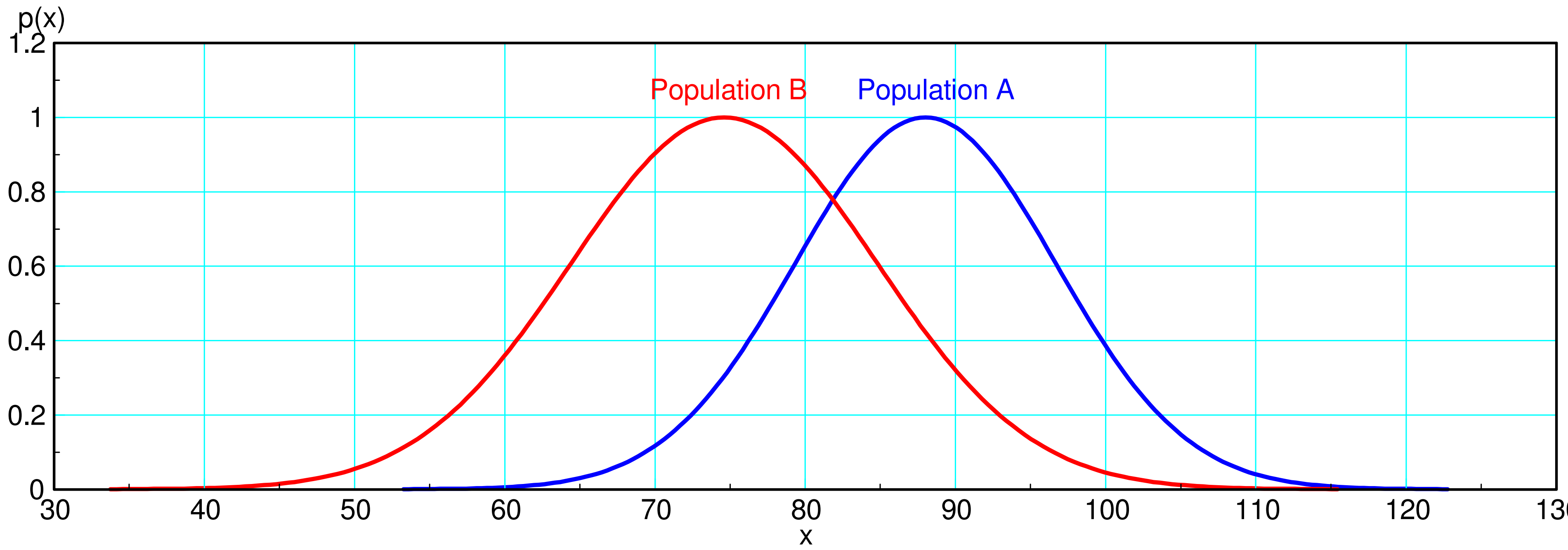
- Gain of a 3904 transistor,
- June temperature in Fargo,
- Player A's score in Hungry-Hungry Hippo



What if you have two populations?

Examples:

- June vs. July: Which month is hotter?
- Player A vs. Player B: Who is going to win?
- Batch A vs. Batch B: Which has a higher gain?



Solution: Force the problem to fit the solution

Create a new variable (you now have a single population)

W = A - B

W has a Student-t distribution

- Central Limit Theorem

Parameters:

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

$$s_w^2 = s_a^2 + s_b^2 \qquad \text{individual}$$

$$s_w^2 = \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b} \qquad \text{population}$$

$$d.f. = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}} \approx \min\left(df_a, df_b\right)$$

Individual vs. Population

Individual:

- If you play one more game, who is going to win?
- Sample size doesn't affect the odds

Population

- Which population has the higher average?
- If you plan an infinite series, who is going to win?
- Sample size does affect the odds
- With a large enough sample, you can identify very small differences

Examples:

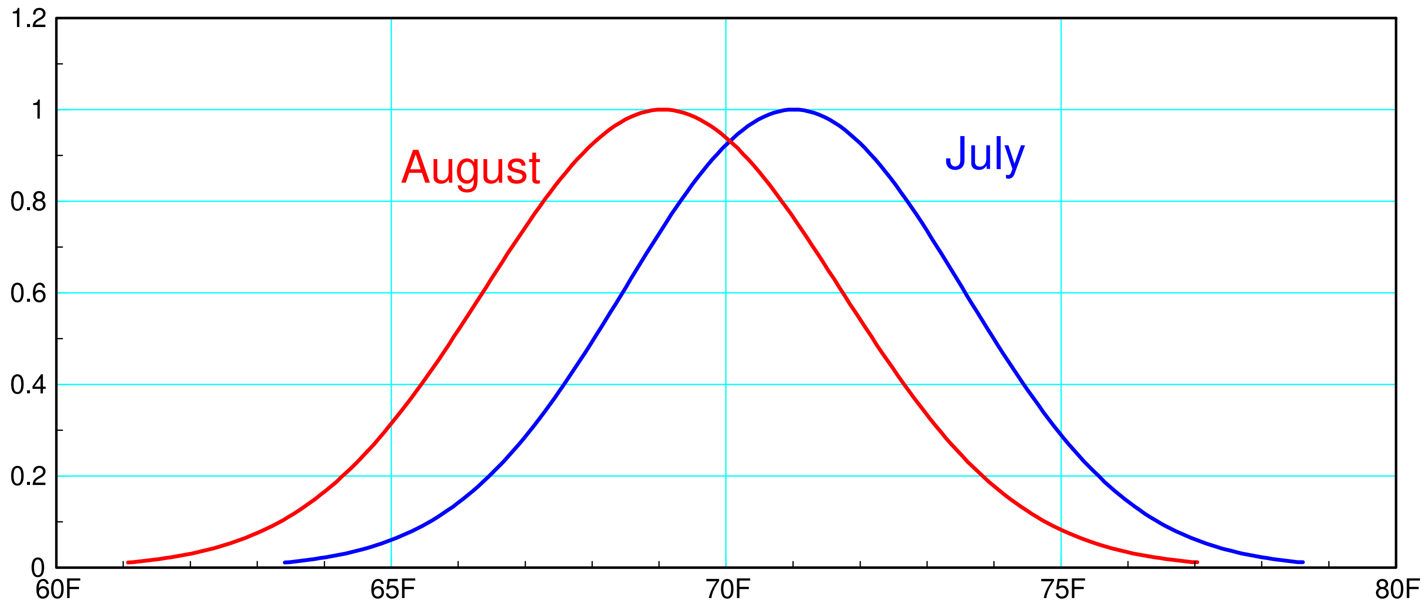
- Fargo temperature: Is July warmer than August?
- Fargo temperature: Was 2000-2020 warmer than 1950-1970?
- Transistor Gains: Does batch A have a higher gain than batch B?
- Reflex Times: Do I have faster reflexes in the morning or afternoon?

Fargo Temperature: July vs. August

Hector Airport has been recording Fargo temperatures since 1942.

Average temperatures for July and August data are:

Pop	Month	Mean	St Dev	n
A	July	71.0107F	2.5337F	81
B	August	69.0522F	2.6617F	81



Will July be warmer in August this year?

Individual Test

Create a new variable, W = A - B

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

$$s_w^2 = s_a^2 + s_b^2$$

Pop	Month	Mean	St Dev	dof
A	July	71.0107F	2.5337F	80
B	August	69.0522F	2.6617F	80
W A - B		1.9585F	3.6748F	min(80, 80)

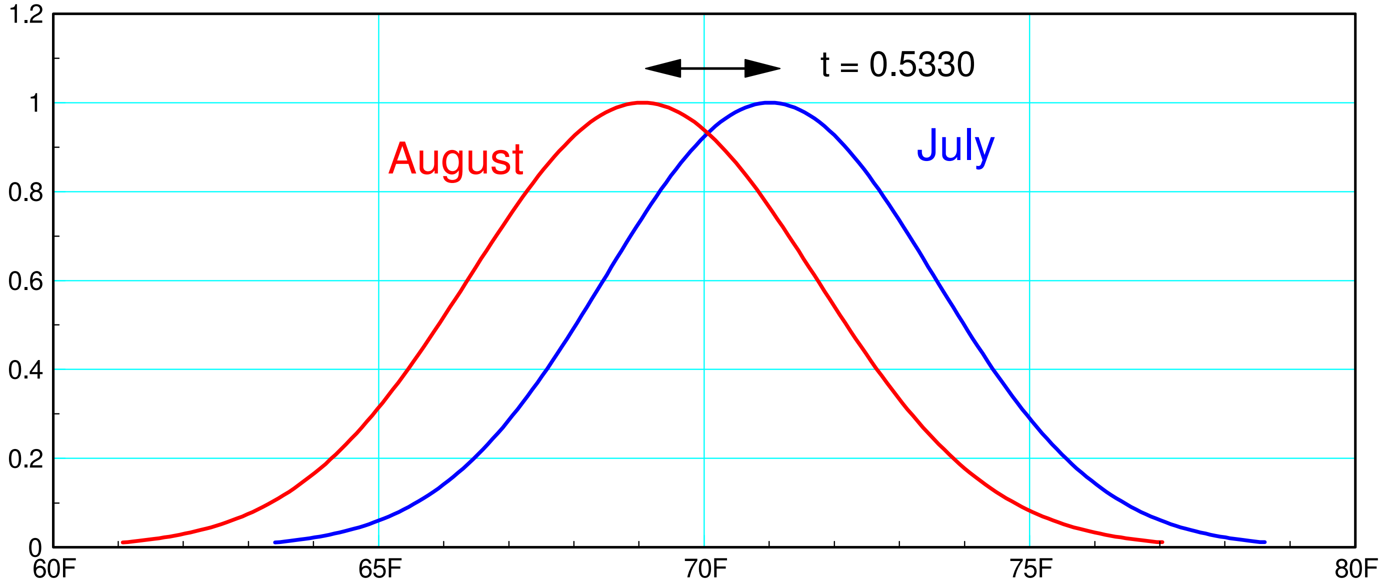
Compute the t-score

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{s_a^2 + s_b^2}}\right) = 0.5330$$

Convert to a probability using a Student-t Table

- 80 degrees of freedom
- p = 70.224%

It is 70.224% likely that July will be hotter than August this year



Which is the hotter month: July or August?

This is a population question

- Which population has the higher overall average

The variance of a population drops with the sample size

- You know more about populations than individuals

$$s_w^2 = \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}$$

Pop	Month	Mean	St Dev	dof
A	July	71.0107F	2.5337F	80
B	August	69.0522F	2.6617F	80
W A - B		1.9585F	0.4083F	min(80, 80)

Compute the t-score

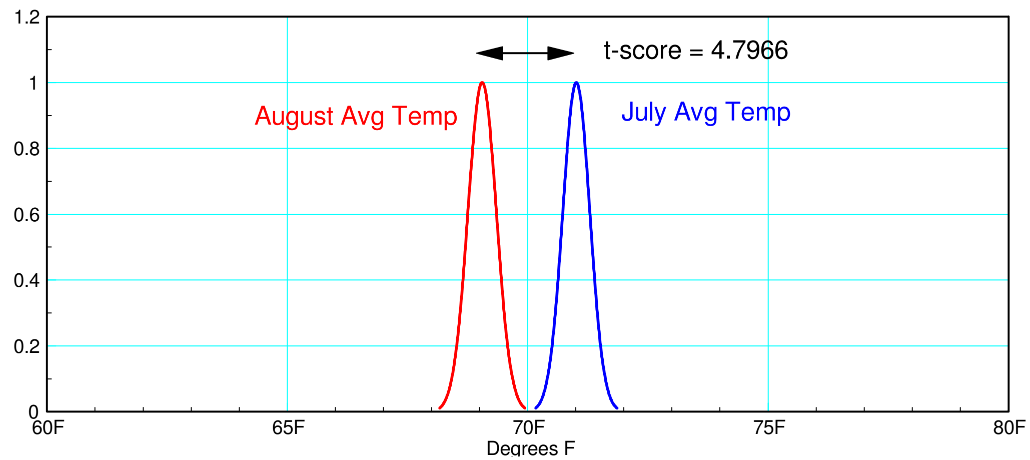
$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}\right) = 4.7966$$

Convert to a probability using a Student-t Table

- 80 degrees of freedom
- $p > 99.999\%$

It is over 99.999% likely that July is hotter overall than August

- You know more about populations than individuals

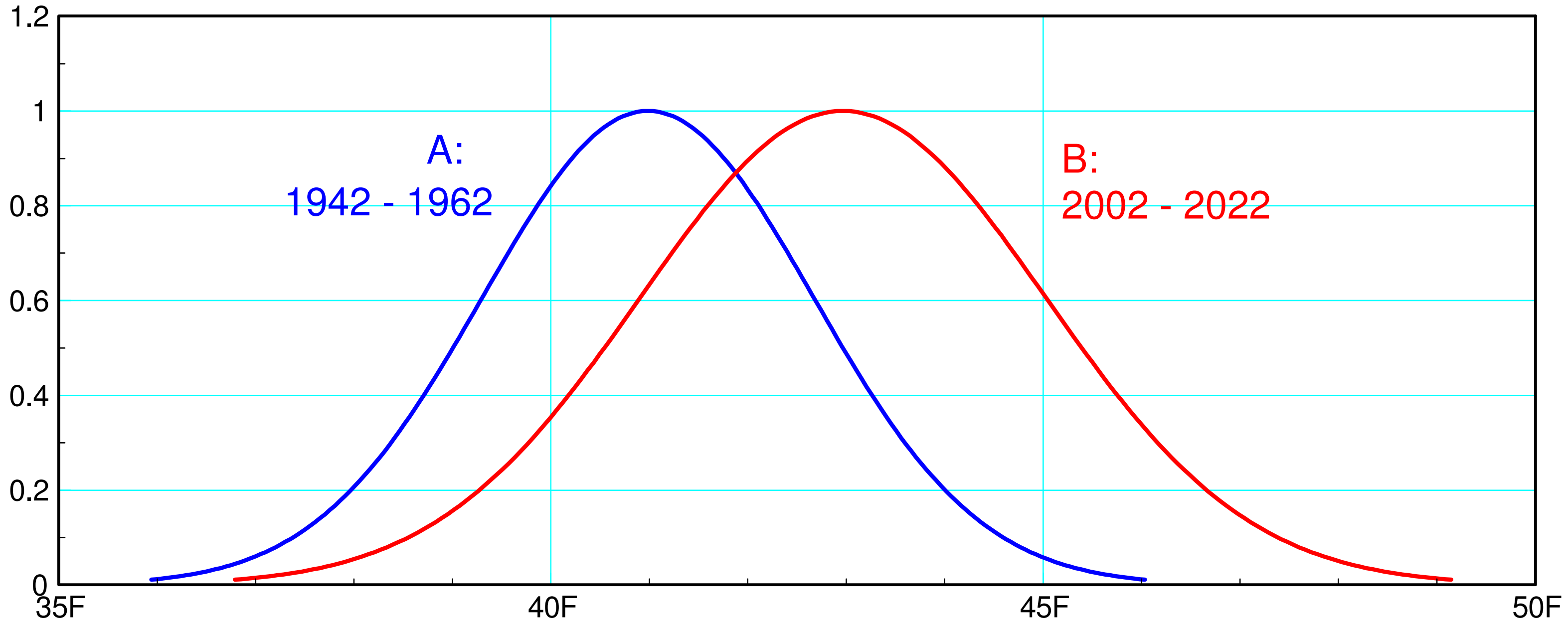


Is Fargo Getting Warmer?

Compare the yearly average temperature in Fargo

source: Hector Airport

Pop	year	Mean	St Dev	dof
A	1942 - 1962	40.9857F	1.6823F	20
B	2002 - 2022	42.9671F	2.0594F	20



Individual Test

- If you pick a random year from A and B, what is the probability that A > B?

Form a new variable W = A - B

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

$$s_w^2 = s_a^2 + s_b^2$$

Pop	Month	Mean	St Dev	dof
A	1942 - 1962	40.9857F	1.6823F	20
B	2002 - 2022	42.9671F	2.0594F	20
W	A - B	-1.9814F	2.6592F	approx 20

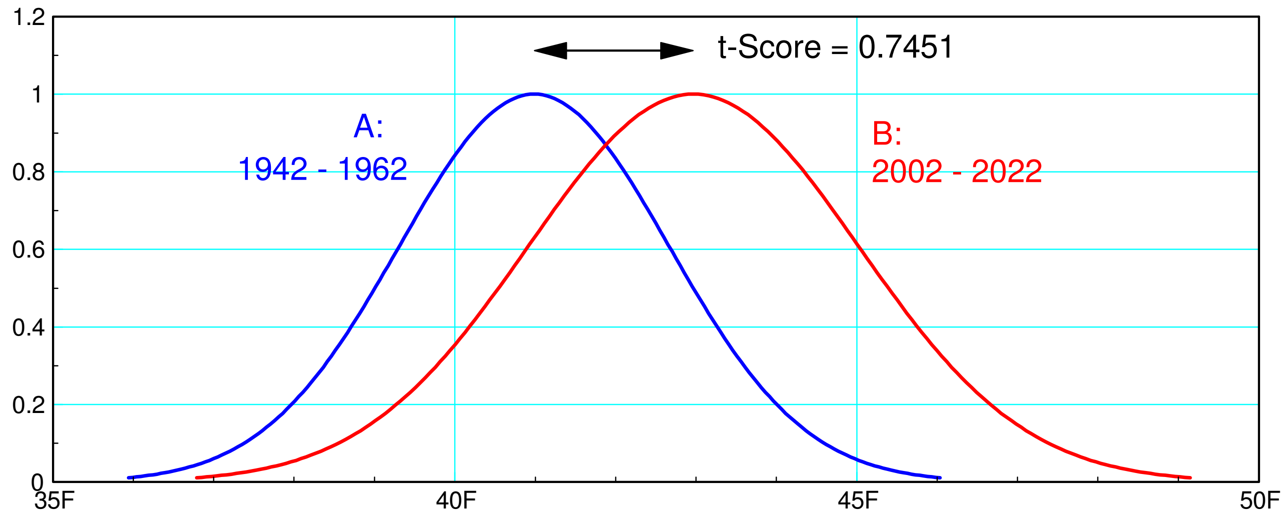
Compute the t-score

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{s_a^2 + s_b^2}}\right) = -0.7451$$

Convert to a probability using a Student-t Table

- 20 degrees of freedom
- p = 0.23244

There is a 23.244% chance than a random year in group A will be warmer than a random year in group B



Population Test

- Is the average temperature over the last 20 years higher than it was 80 years ago?

The variance drops with the sample size

$$s_w^2 = \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}$$

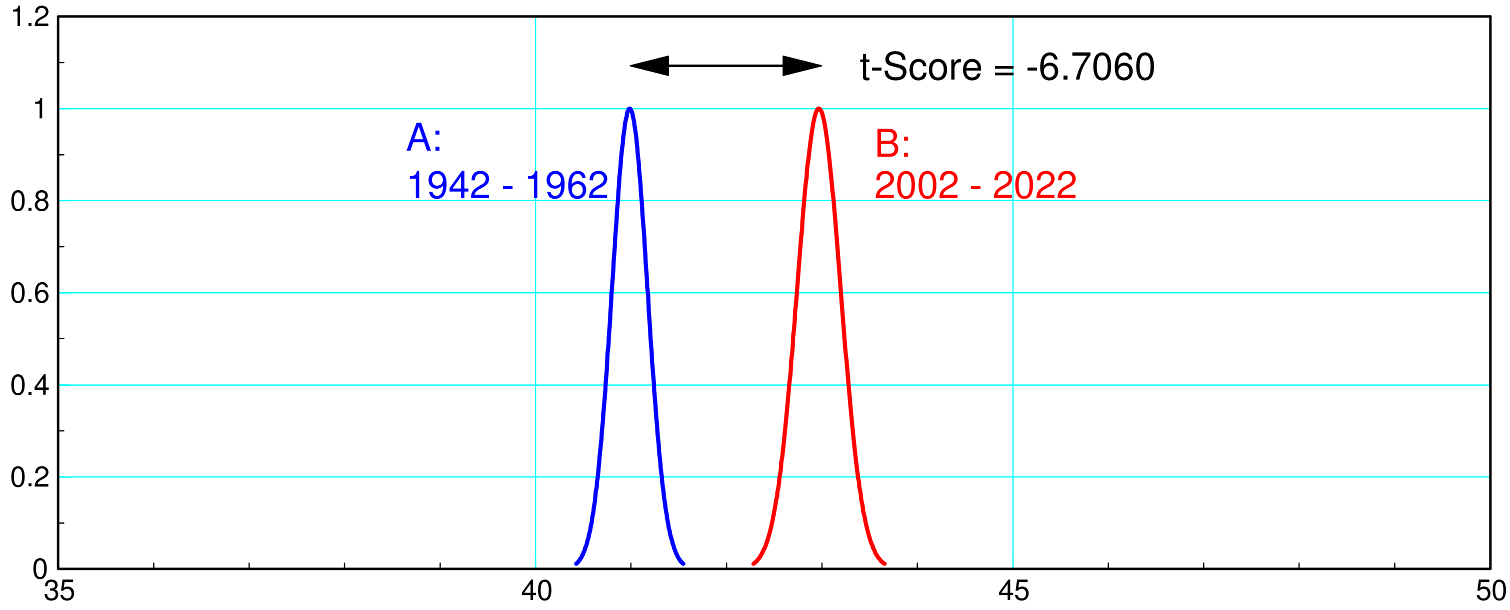
Pop	Month	Mean	St Dev	dof
A	1942 - 1962	40.9857F	1.6823F	20
B	2002 - 2022	42.9671F	2.0594F	20
W	A - B	-1.9814F	0.2955F	approx 20

The t-score increases with the sample size

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}\right) = -6.7060$$

Use a t-Table to convert this to a probability

- p > 99.999%
- It is more than 99.999% certain that the last 20 years were warmer



Fun with t-Tests

With t-tests, you can start asking some interesting questions...

- Are polyester film capacitors polarized?
- Are electrolytic capacitors polarized?
- Does temperature affect the capacitance of polyester film capacitors?

Other things you could test...

- Do loud noise affect my reaction time?
- Is my reaction time affected by two drinks?
- Does using alcohol-free gas improve my gas mileage?
- Does adding a lid to a coffee cup help keep the contents warm?
- Does adding a spoon to a cup of coffee make it cool off faster?

Are Polyester Film Capacitors Polarized?

Procedure: Measure the capacitance of ten polyester film capacitors

- +/- going left to right
- +/- going right to left

Data:

Polarity	mean (nF)	st dev (nF)	n
A: Correct	384.3280	3.8993	10
B: Reverse	384.4410	3.9067	10
W = A - B	-0.1130	5.5197 individual 1.7455 population	10

Analysis

Form the t-score

Individual (9 dof)

$$t = \left(\frac{-0.1130-0}{5.5107}\right) = -0.02047$$

$$p = 49.2\%$$

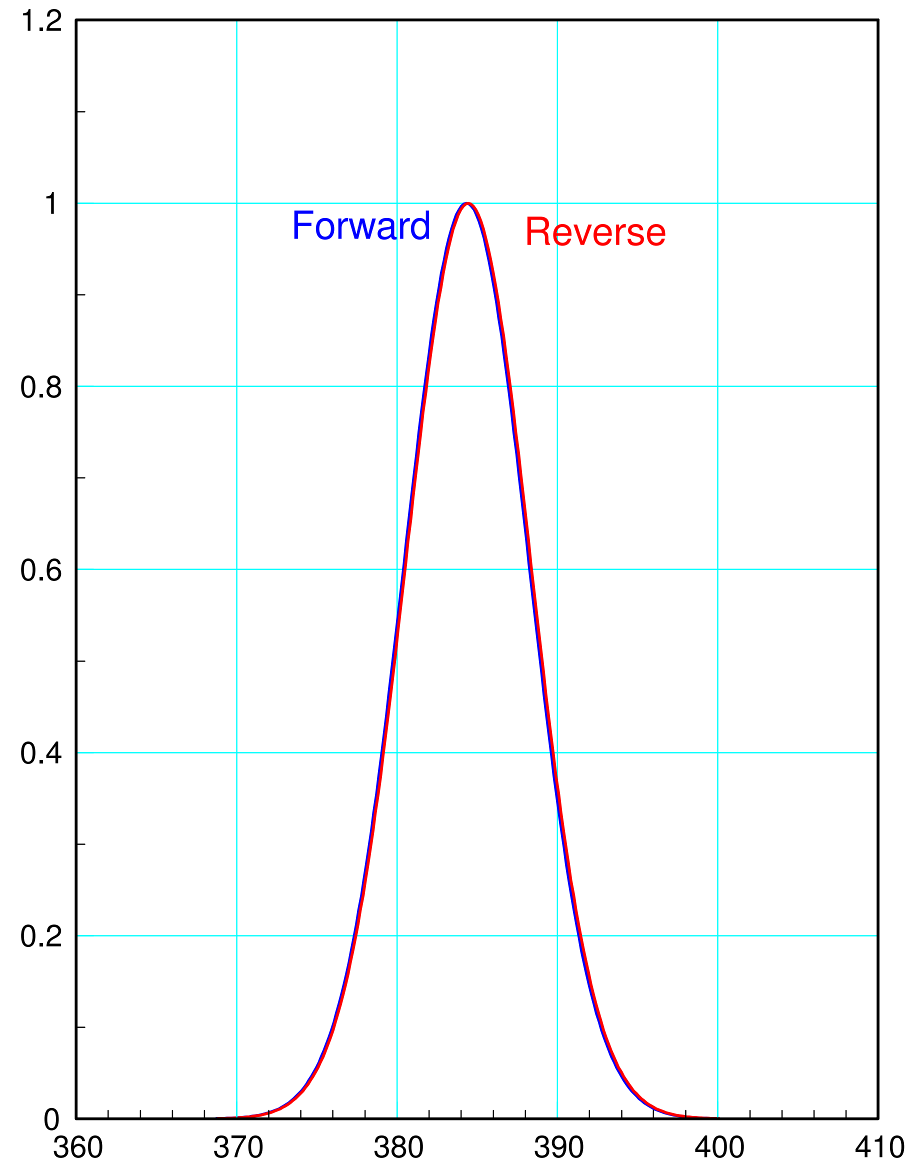
Population (9 dof)

$$t = \left(\frac{-0.1130}{1.7455}\right) = -0.06474$$

$$p = 47.49\%$$

Result: Polarity doesn't make much difference.

- The polarity with the higher capacitance is essentially a 50/50 split



Polarity of Polyester Film Capacitors (take 2)

- If you have the raw data, you can compute the difference for each capacitor

	A (nF)	B (nF)	C = A - B
1	387.82	387.97	-0.15
2	382.30	382.40	-0.10
3	385.80	385.85	-0.05
4	384.53	384.52	0.01
5	383.08	383.20	-0.12
6	383.69	383.81	-0.12
7	377.67	377.80	-0.13
8	385.40	385.57	-0.17
9	380.95	381.09	-0.14
10	392.04	392.20	-0.16
x	384.3280	384.4410	-0.1130
s	3.8993	3.9067	0.0550

Analysis:

Individual:

$$t = \left(\frac{-0.1130}{0.0550}\right) = -2.0545$$

$$p = 0.03505$$

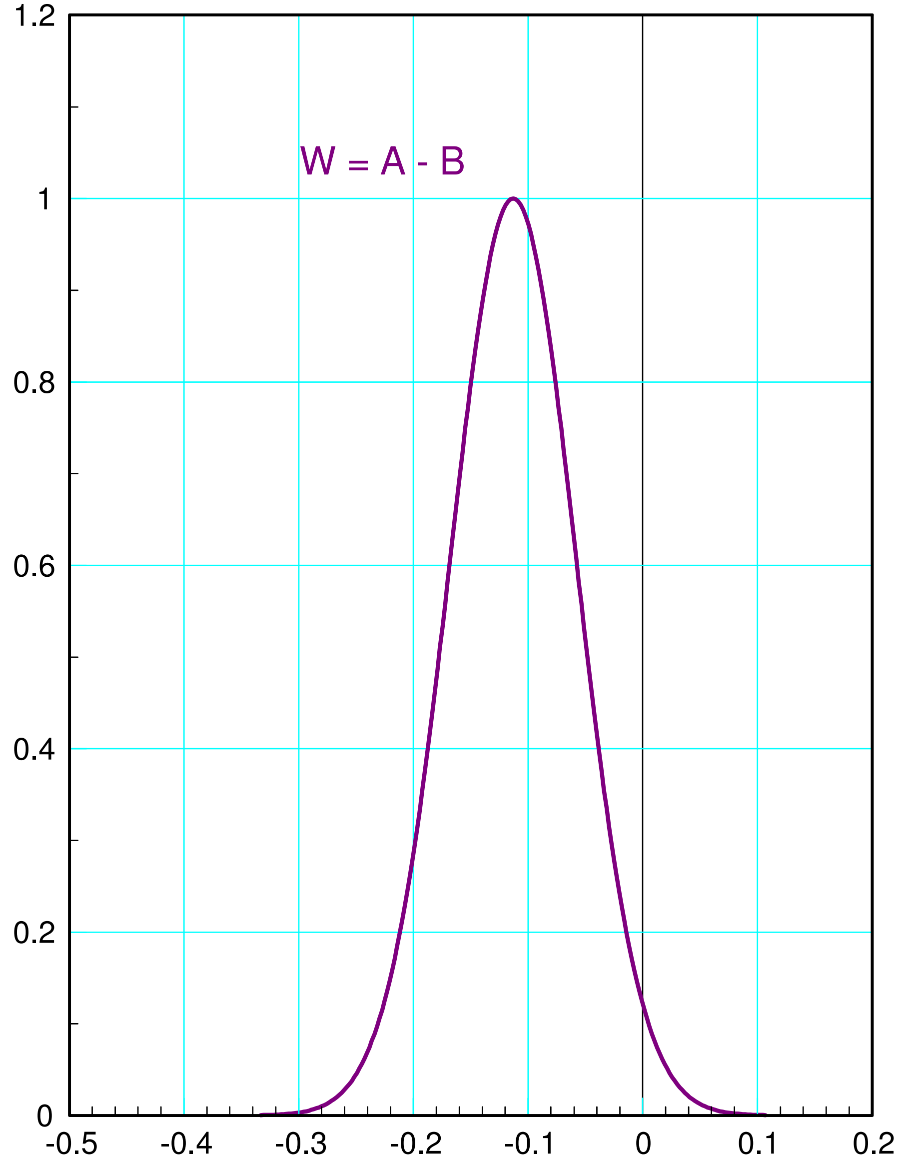
Population:

$$t = \left(\frac{-0.1130}{\left(\frac{0.0550}{\sqrt{10}}\right)}\right) = -6.4970$$

$$p = 0.00006$$

I'm 99.994% certain that polarity affects the polarity of polyester film capacitors

- The difference may be 0.04%, but it's there and it's measurable



Polarity of Electrolytic Capacitors

- Does polarity matter on a electrolytic capacitor?

	A (nF)	B (nF)	C = A - B
1	997.54	997.76	-0.22
2	1024.26	1025.30	-1.04
3	990.72	991.74	-1.02
4	1016.34	1017.96	-1.62
5	1003.76	1006.16	-2.40
6	1013.32	1015.86	-2.54
7	992.8	993.76	-0.96
8	1005.36	1006.50	-1.14
9	1004.36	1005.56	-1.20
10	991.78	993.50	-1.72
x	1004.02	1005.41	-1.386
s	11.277	11.485	0.701

If the only data you have is the mean and standard deviation...

W = A - B (individual)

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = -1.39nF$$

$$s_w = \sqrt{s_a^2 + s_b^2} = 16.09nF$$

$$t = \left(\frac{-1.39nF}{16.09nF}\right) = -0.0864$$

p = 0.46652

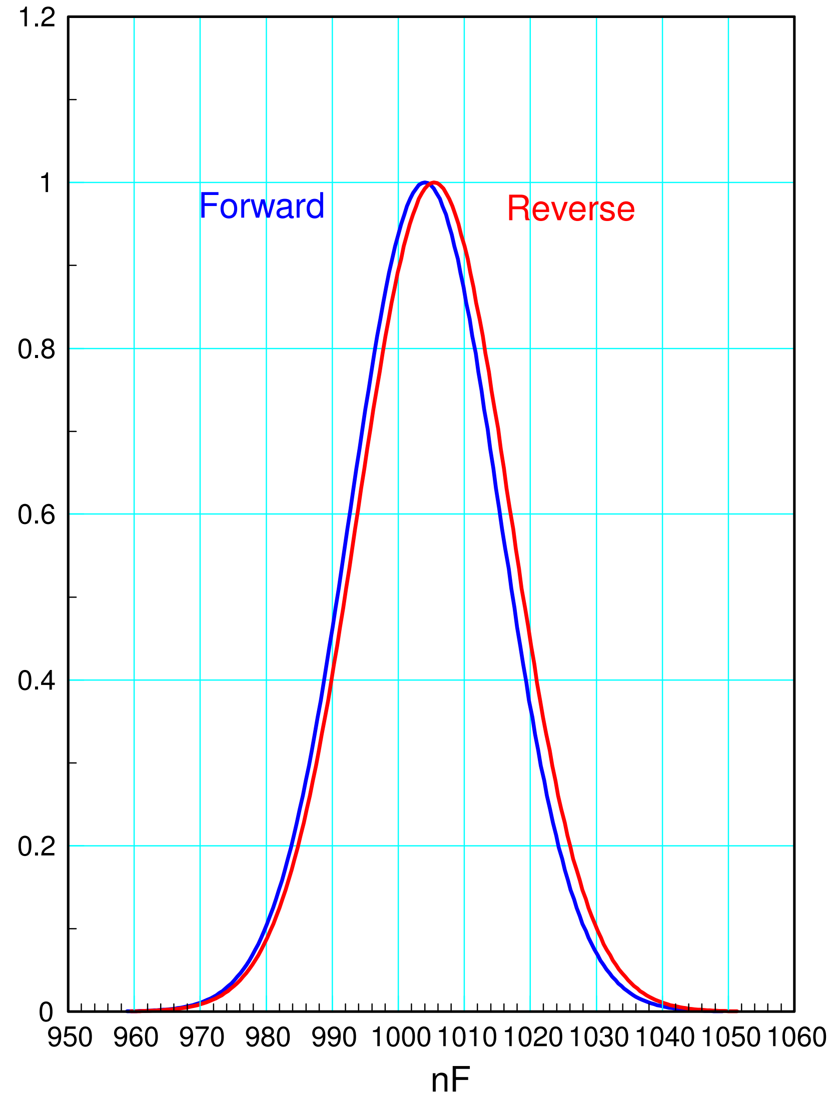
Population

$$s_w = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}} = 5.0900nF$$

$$t = \left(\frac{-1.39nF}{5.0900nF}\right) = -0.2371$$

p = 0.40894

Polarity doesn't have a significant affect on capacitance



If you keep track of each capacitor's value

- W = difference between capacitance

Individual Capacitor

$$\bar{x}_w = -1.386nF$$

$$s_w = 0.701$$

$$t = \left(\frac{-1.386nF}{0.701nF}\right) = -1.9772$$

p = 0.03971

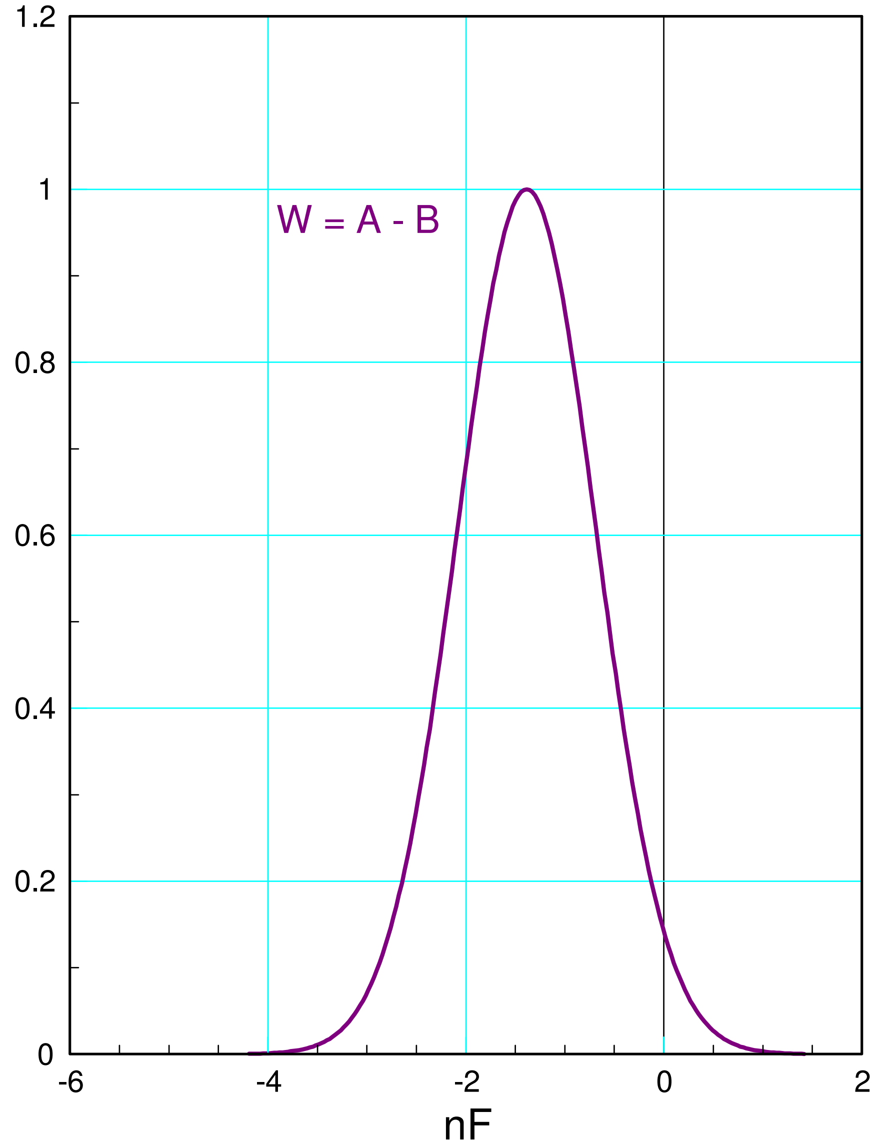
Population

$$s_w = \left(\frac{0.701}{\sqrt{10}}\right)$$

$$t = -6.2524$$

p = 0.00007

There is a statistically significant difference in capacitance when you switch the polarity



Does Temperature Affect Polyester Film Capacitors?

	A: +83F (nF)	B: +40F (nF)	C = A - B (nF)
1	400.59	398.35	2.24
2	406.15	403.80	2.35
3	406.34	403.76	2.58
4	398.45	396.51	1.94
5	400.75	398.46	2.29
6	403.92	401.54	2.38
7	404.35	402.51	1.84
8	410.59	408.37	2.22
9	413.8	411.52	2.28
10	405.46	402.93	2.53
11	409.73	407.78	1.95
x	405.466	403.230	2.2364
s	4.6260	4.6005	0.2387

If all you have is the mean, standard deviation, and sample size...

(population analysis)

$$W = A - B$$

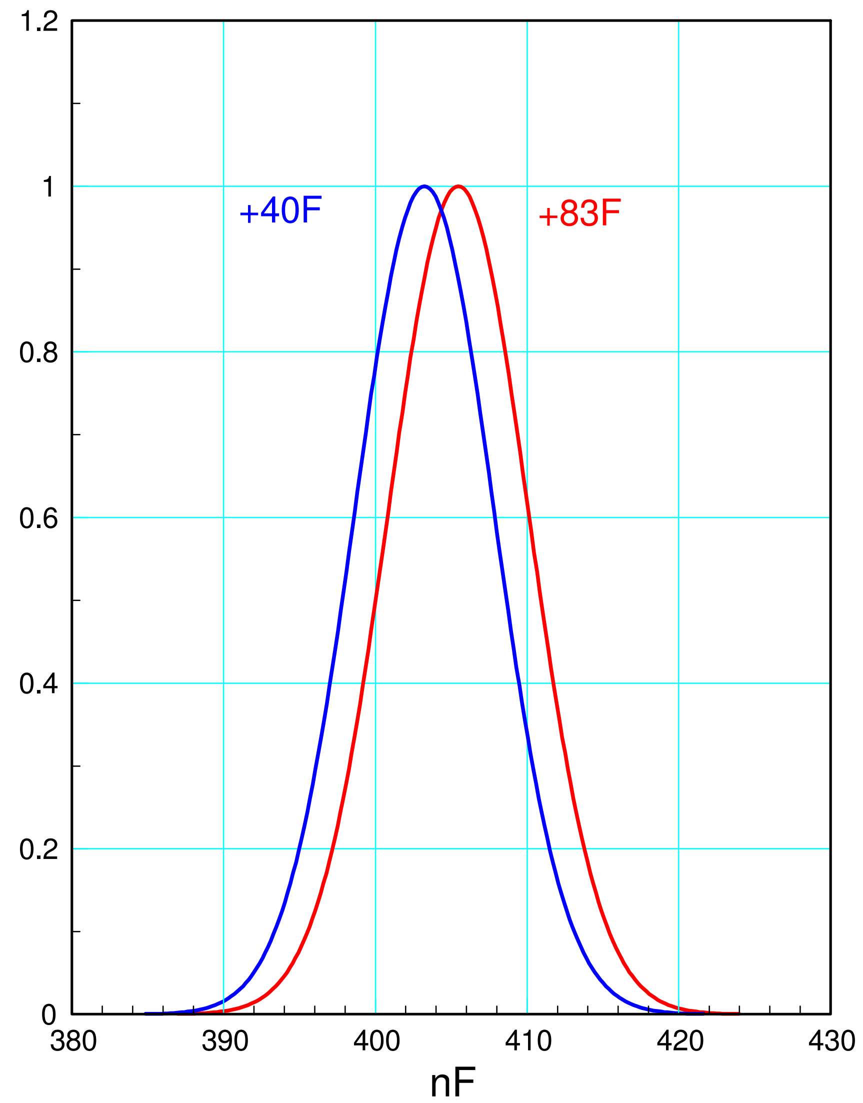
$$\bar{x}_w = \bar{x}_a - \bar{x}_b = 2.236nF$$

$$s_w = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}} = 1.9671nF$$

$$t = \frac{\bar{x}_w}{s_w} = 1.1367$$

p = 0.85892

It's 85% likely that capacitance drops with temperature



If you keep track of the value of each capacitor

- at 83F
- at 40F

$$W = A - B$$

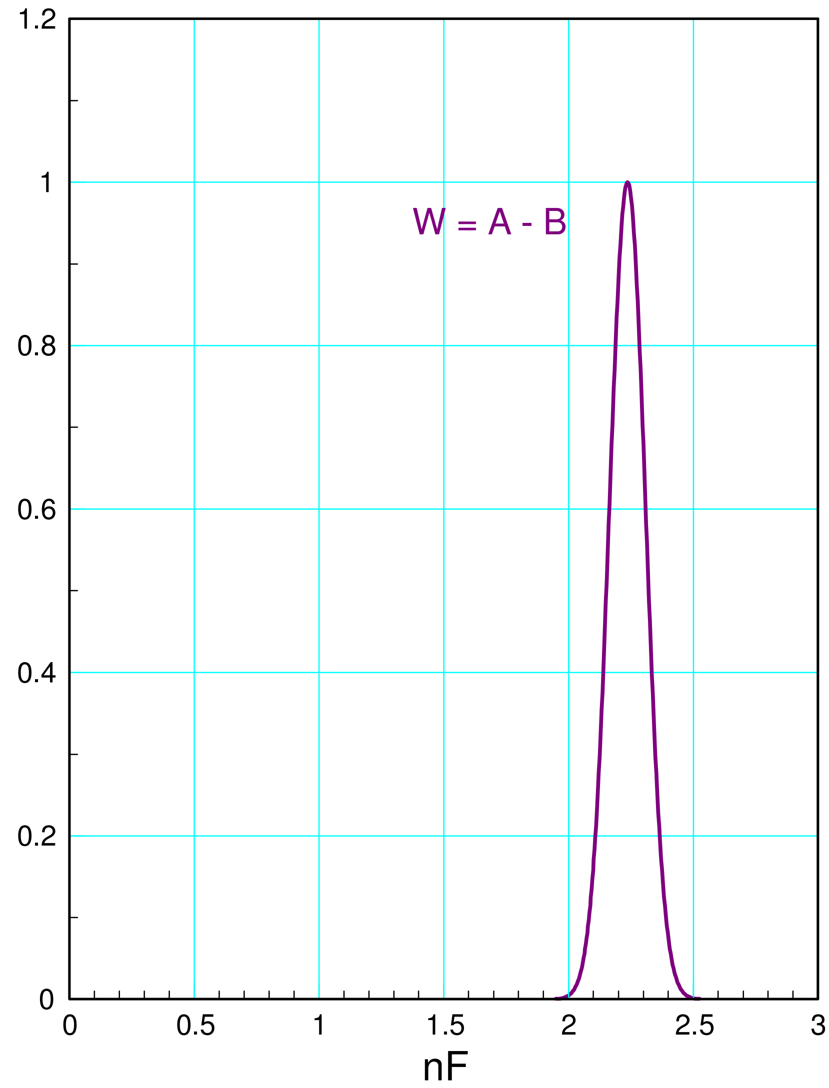
$$x_w = 2.2364nF$$

$$s_w = \sqrt{\frac{0.2387^2}{11}} = 0.0720nF$$

$$t = \frac{x_w}{s_w} = 31.0737$$

$$\text{p} > 0.99999$$

The temperature affect on a polyester film capacitor is statistically significant at A 99.999% confidence level



Summary

With a t-test, you can compare two populations

- Create a new variable, W = A - B
- Determine the probability that W > 0

Keep all of your data

- At a minimum, keep the mean, standard deviation, and sample size
- There may be ways to pull information out if you keep all of your data

Only really works with two populations

- If you have more than two populations, you need a different tool
- ANOVA is one such tool (upcoming....)