

“Most of the global average warming over the past 50 years is *very likely* due to anthropogenic GHG increases...”

How does the IPCC know whether the statement about global warming’s cause is *extremely likely*, *very likely*, or *more likely than not*? How does the IPCC consistently, across many different disciplines (scientific, technical, and socio-economic), quantify uncertainty? The IPCC uses three different approaches for uncertainty, depending on whether the data are qualitative, quantitative, or based on expert judgment. For the quantitative assessments, most often used in the scientific disciplines, the IPCC uses a Likelihood Scale to consistently define the probability, or likelihood of occurrence. The Likelihood Scale is based on statistics and probability. Statistics is the language that scientists generally use to objectively, and consistently make conclusions about their data – despite uncertainty that is inherent in all datasets. However, statistics is not the best way to communicate findings to a broad audience for the purpose of making new policies or educating the public. By using the Likelihood Scale, the IPCC can effectively communicate what we know and what we don’t know about global climate change.

In this exercise, you will use statistics to analyze a dataset from Lake Mendota that spans the last 150 years. Every year, since 1855, someone has recorded when the lake froze (ice on), and when the lake thawed (ice off). We are going to use these data to ask, ‘Is ice off date on Lake Mendota earlier?’ To answer this question and at the same time quantify the uncertainty around the answer, you will use one of the most basic statistical techniques – a t-test.



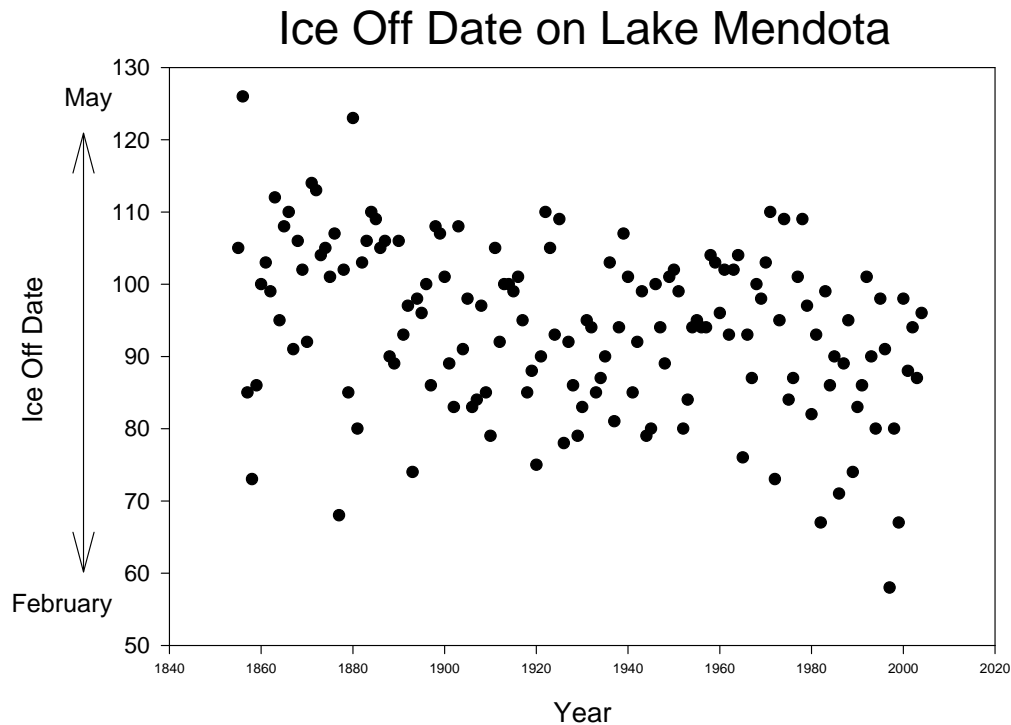
The learning objectives for this lab are to:

- 1) Analyze a historical, climatic dataset *qualitatively* and then *quantitatively*, using a probabilistic technique – the t-test.
- 2) Relate statistical results to the IPCC Likelihood Scale.
- 3) Consider how statistics provide a consistent, objective way of making conclusions despite uncertainty in scientific data.

## Qualitative analysis

Historical climate data show a substantial amount of year-to-year variation. This variation is one form of uncertainty when scientists analyze for trends in climate data. For example, if you were alive in 1920 and you looked back at the data would you say, with certainty, ice off date is earlier? Think about how you determine your level of confidence – especially considering the year-to-year variation.

- Study the graph below to make a *qualitative* assessment of ice off date. Record whether ice off date today is earlier than it was in the 1850's. Report your level of certainty (using the language from the IPCC Likelihood Scale).



Is ice off date earlier today?	
Level of certainty	

## Quantitative analysis

Without quantitative, probabilistic analysis, it is difficult to analyze data with certainty. Now that you have qualitatively analyzed Mendota's ice off data, you will *quantitatively* analyze the same data. The dataset is provided as an excel file, and is divided into several twenty year sections. You are going to ask, using statistics, 'Is ice off date on Lake Mendota earlier?'. First, you will ask the question as if you were alive in 1914 (Is the first twenty year section statistically different from the second twenty year section), and then again in 1934, and so on until 1994. In this way, you will be able to visualize how the level of certainty changes as time goes on.

You will use a very basic statistical test, the t-test to answer these questions. A t-test compares means between two samples, while also considering the variation around each mean. You will use Excel to run the t-test, but it is important to note the actual equation Excel is using:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \text{ where } s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2 + s_2^2}{n}}$$

Where  $\bar{X}$  = sample mean and  $s$  = standard deviation, and  $n$  = sample size.

## Statistical analysis

1. Open the Excel spreadsheet **Ice\_Off\_Date.xls**
2. Click on **Tools** and then click on **Data Analysis**. (Note: If you don't see **Data Analysis**, then select **Add-ins** under the Tools tab, and click on the **Analysis ToolPak** the click **OK**. Now go back to click on **Tools** and you should see **Data Analysis**).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Year	Ice off Date Section 1	Year	Ice off Date Section 2	Year	Ice off Date Section 3	Year	Ice off Date Section 4	Year	Ice off Date Section 5	Year	Ice off Date Section 6	Year	Ice off Date Section 7	
1	1855	105	1875	101	1895	96	1915	99	1935	90	1955	95	1975	84	
2	1856	126	1876	107	1896	100	1916	101	1936	103	1956	94	1976	87	
3	1857	85	1877	68	1897	86	1917	95	1937	81	1957	94	1977	101	
4	1858	73	1878	102	1898	108	1918	85	1938	94	1958	104	1978	109	
5	1859	86	1879	85	1899	107	1919	88	1939	107	1959	103	1979	97	
6	1860	100	1880	123	1900	101	1920	75	1940	101	1960	96	1980	82	
7	1861	103	1881	80	1901	89	1921	90	1941	85	1961	102	1981	93	
8	1862	99	1882	103	1902	83	1922	110	1942	92	1962	93	1982	67	
9	1863	112	1883	106	1903	108	1923	105	1943	99	1963	102	1983	99	
10	1864	95	1884	110	1904	91	1924	93	1944	79	1964	104	1984	86	
11	1865	108	1885	109	1905	98	1925	109	1945	80	1965	76	1985	90	
12	1866	110	1886	105	1906	83	1926	78	1946	100	1966	93	1986	71	
13	1867	91	1887	106	1907	84	1927	92	1947	94	1967	87	1987	89	
14	1868	106	1888	90	1908	97	1928	86	1948	89	1968	100	1988	95	
15	1869	102	1889	89	1909	85	1929	79	1949	101	1969	98	1989	74	
16	1870	92	1890	106	1910	79	1930	83	1950	102	1970	103	1990	83	
17	1871	114	1891	93	1911	105	1931	95	1951	99	1971	110	1991	86	
18	1872	113	1892	97	1912	92	1932	94	1952	80	1972	73	1992	101	
19	1873	104	1893	74	1913	100	1933	85	1953	84	1973	95	1993	90	
20	1874	105	1894	98	1914	100	1934	87	1954	94	1974	109	1994	80	

3. Scroll down towards the bottom and select **t-Test: Two-sample Assuming Equal Variances**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Year	Ice off Date Section 1	Year	Ice off Date Section 2	Year	Ice off Date Section 3	Year	Ice off Date Section 4	Year	Ice off Date Section 5	Year	Ice off Date Section 6	Year	Ice off Date Section 7	
1	1855	105	1875	101	1895	96	1915	99	1935	90	1955	95	1975	84	
2	1856	126	1876	107	1896	100	1916	101	1936	103	1956	94	1976	87	
3	1857	85	1877	68	1897	86	1917	95	1937	81	1957	94	1977	101	
4	1858	73	1878	102	1898	108	1918	85	1938	94	1958	104	1978	109	
5	1859	86	1879	85	1899	107	1919	88	1939	107	1959	103	1979	97	
6	1860	100	1880	123	1900	101	1920	75	1940	101	1960	96	1980	82	
7	1861	103	1881	80	1901	89	1921	90	1941	85	1961	102	1981	93	
8	1862	99	1882	103	1902	83	1922	110	1942	92	1962	93	1982	67	
9	1863	112	1883	106	1903	108	1923	105	1943	99	1963	102	1983	99	
10	1864	95	1884	110	1904	91	1924	93	1944	79	1964	104	1984	86	
11	1865	108	1885	109	1905	98	1925	109	1945	80	1965	76	1985	90	
12	1866	110	1886	105	1906	83	1926	78	1946	100	1966	93	1986	71	
13	1867	91	1887	106	1907	84	1927	92	1947	94	1967	87	1987	89	
14	1868	106	1888	90	1908	97	1928	86	1948	89	1968	100	1988	95	
15	1869	102	1889	89	1909	85	1929	79	1949	101	1969	98	1989	74	
16	1870	92	1890	106	1910	79	1930	83	1950	102	1970	103	1990	83	
17	1871	114	1891	93	1911	105	1931	95	1951	99	1971	110	1991	86	
18	1872	113	1892	97	1912	92	1932	94	1952	80	1972	73	1992	101	
19	1873	104	1893	74	1913	100	1933	85	1953	84	1973	95	1993	90	
20	1874	105	1894	98	1914	100	1934	87	1954	94	1974	109	1994	80	

4. A dialogue box will open with Input and Output options. Click on the first box for **Variable 1 Range**. Now move your cursor over to the first entry for 'Ice Off Date Section 1' (cell B2), left click and drag to the last entry for the column (cell B21). Now click on the box for **Variable 2 Range**. Select the data in 'Ice Off Date Section 2' column (cells D2:D21).
5. Now enter 0 in the box for **Hypothesized mean difference**.
6. Keep the **labels** box unchecked.
7. Keep the **alpha** level at 0.05.
8. In the output section, select **New Worksheet Ply** and enter 1 vs. 2 in the box.
9. Click **OK**.

The screenshot shows a Microsoft Excel spreadsheet with a data table and a 't-Test: Two-Sample Assuming Equal Variances' dialog box. The data table has columns for Year and Ice off Date, grouped into seven sections. The dialog box is open, showing the input ranges for Variable 1 (1852:1862) and Variable 2 (1892:1902), with a hypothesized mean difference of 0 and an alpha level of 0.05.

Year	Ice off Date	Year	Ice off Date	Year	Ice off Date	Year	Ice off Date	Year	Ice off Date	Year	Ice off Date	Year	Ice off Date
1855	105	1875	101	1895	95	1915	99	1935	90	1955	95	1975	84
1856	126	1876	107	1896	100	1916	101	1936	103	1956	94	1976	87
1857	85	1877	68	1897	86							1977	101
1858	73	1878	102	1898	108							1978	109
1859	86	1879	85	1899	107							1979	97
1860	100	1880	123	1900	101							1980	82
1861	103	1881	80	1901	89							1981	93
1862	99	1882	103	1902	83							1982	67
1863	112	1883	106	1903	108							1983	99
1864	95	1884	110	1904	91							1984	86
1865	108	1885	109	1905	98							1985	90
1866	110	1886	105	1906	83							1986	71
1867	91	1887	106	1907	84							1987	69
1868	106	1888	90	1908	97							1988	95
1869	102	1889	89	1909	85							1989	74
1870	92	1890	106	1910	79							1990	83
1871	114	1891	93	1911	105							1991	86
1872	113	1892	97	1912	92	1932	94	1952	80	1972	73	1992	101
1873	104	1893	74	1913	100	1933	85	1953	84	1973	95	1993	90
1874	105	1894	98	1914	100	1934	87	1954	94	1974	109	1994	80

Excel has calculated a t-statistic to determine whether mean ice off date has changed on Lake Mendota from the time period in section 1 (1855-1874) compared to the time period in section 2 (1875-1894). The results are in a new sheet labeled **1 vs. 2** and should look like this:

The screenshot shows a new Excel sheet titled '1 vs. 2' with a table of t-test results. The t-statistic is displayed in cell B10 as 0.957649.

	Variable 1	Variable 2
Mean	101.45	97.6
Variance	144.3658	178.8842
Observations	20	20
Pooled Variance	161.625	
Hypothesized Mean Difference	0	
df	38	
t Stat	0.957649	
P(T<=t) one-tail	0.172147	
t Critical one-tail	1.685954	
P(T<=t) two-tail	0.344293	
t Critical two-tail	2.024394	

The **t Stat**, in cell B10 is the number we are interested in analyzing. Statisticians use a look-up table to determine whether a particular t-statistic, with a particular sample size, shows that two samples actually differ. The IPCC has basically taken this same look-up table and added their own language to describe the likelihood of various climatic changes in a consistent way.

t-statistic	Probability of occurrence	IPCC Likelihood Scale
0.681	75%	Likely
0.851	80%	
1.05	85%	
1.303	90%	Very Likely
1.684	95%	Extremely Likely
2.021	97.5%	
2.123	98%	
2.423	99%	Virtually Certain
2.704	99.5%	
2.971	99.75%	
3.307	99.9%	
3.551	99.95%	
3.790	99.99%	

The t-statistic for our analysis (0.958) says that the probability that ice off date changed between the two sections of time is somewhere between 80 – 85%. The IPCC would then conclude, “It is *likely* that by the year 1894, ice off date was earlier on Lake Mendota in Wisconsin.”

- Hypothesize how likelihood will change as you move along in time, based on looking at the graph of ice off data.
- Run the same analyses to test for differences in ice off date between section 1 and subsequent sections (1 vs. 3 then 1 vs. 4 and finally 1 vs. 7). When you click on Tools and select Data Analysis, the dialog box should open with your last specifications still in place. All you need to change is the **Variable 2 Range** (F2:F21 for section 3, H2:H21 for section 4, N2:N21 for section 7) and the **New Worksheet Ply** (1 vs. 3, 1 vs. 4, and 1 vs. 7).

The screenshot shows a Microsoft Excel spreadsheet with the following data table:

Year	Ice off Date Section 1	Year	Ice off Date Section 2	Year	Ice off Date Section 3	Year	Ice off Date Section 4	Year	Ice off Date Section 5	Year	Ice off Date Section 6	Year	Ice off Date Section 7	
1855	105	1875	101	1895	96	1915	99	1935	90	1955	95	1975	84	
1856	126	1876	107	1896	100	1916	101	1936	103	1956	94	1976	87	
4	1857	85	1877	68	1897	86						1977	101	
5	1858	73	1878	102	1898	108						1978	109	
6	1859	86	1879	85	1899	107						1979	97	
7	1860	100	1880	123	1900	101						1980	82	
8	1861	103	1881	80	1901	89						1981	93	
9	1862	99	1882	103	1902	83						1982	67	
10	1863	112	1883	106	1903	108						1983	99	
11	1864	95	1884	110	1904	91						1984	86	
12	1865	108	1885	109	1905	98						1985	90	
13	1866	110	1886	105	1906	83						1986	71	
14	1867	91	1887	106	1907	84						1987	89	
15	1868	106	1888	90	1908	97						1988	95	
16	1869	102	1889	89	1909	85						1989	74	
17	1870	92	1890	106	1910	79						1990	83	
18	1871	114	1891	93	1911	105						1991	86	
19	1872	113	1892	97	1912	92	1932	94	1952	80	1972	73	1992	101
20	1873	104	1893	74	1913	100	1933	85	1953	84	1973	95	1993	90
21	1874	105	1894	98	1914	100	1934	87	1954	94	1974	109	1994	80

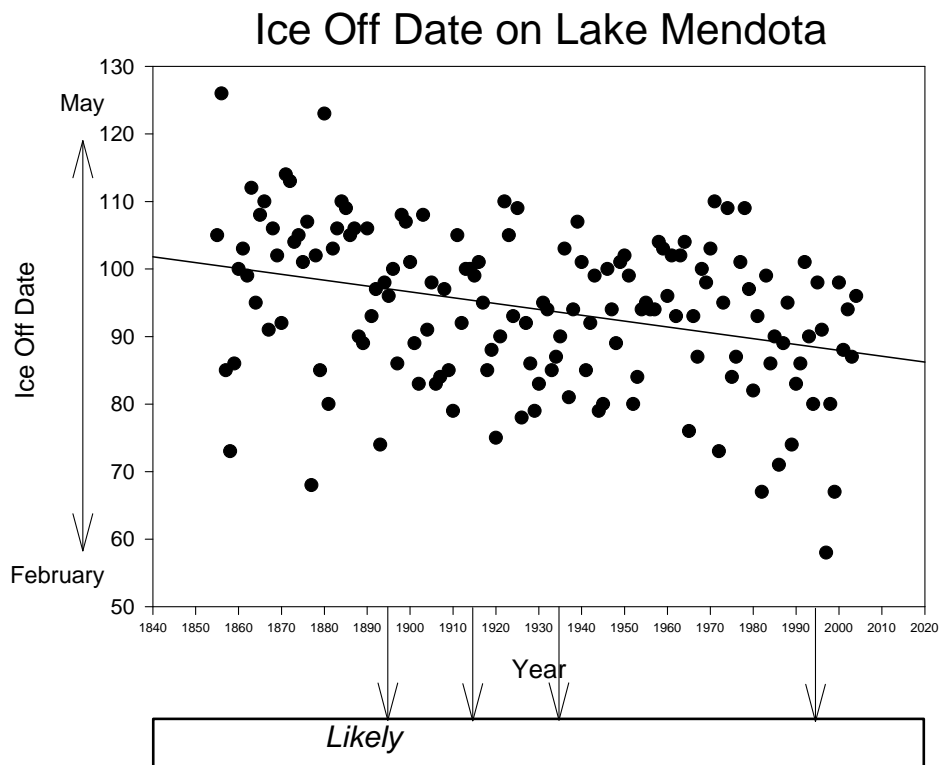
The 't-Test: Two-Sample Assuming Equal Variances' dialog box is open, showing the following settings:

- Variable 1 Range: \$B\$2:\$B\$21
- Variable 2 Range: \$F\$2:\$F\$21
- Hypothesized Mean Difference: 0
- Alpha: 0.05
- Output options:  New Worksheet Ply: 1v3

- Enter your results on a table like this:

Which comparison?	T Stat	Probability of Occurrence	IPCC Likelihood Scale
1 vs. 2	0.958	< 85%	Likely
1 vs. 3			
1 vs. 4			
1 vs. 7			

- Report the Likelihood Scale results on the graph below:



- Did the certainty level for 1994 match the certainty level you reported for the *qualitative* analysis? Why or why not?
- Did the level of certainty change over time as you hypothesized it would?