

Probabilities, Uncertainties & Units Used to Quantify Climate Change

“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.



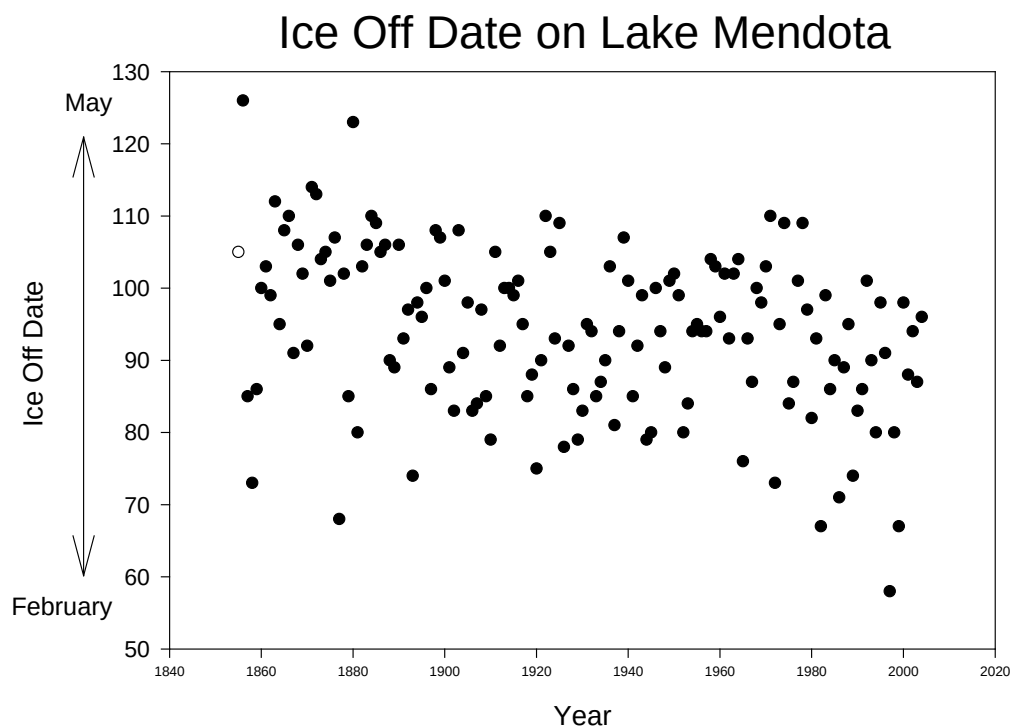
The learning objectives for this lab are to:

- 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 different 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 – (Excel 2007)

1. Open the Excel spreadsheet [150 year dataset from Lake Mendota](#)
2. Click on **Tools** and then click on **More Functions** and **Statistical**. In the Statistical list, scroll down to **TTEST**.

The screenshot shows an Excel spreadsheet with columns for 'Ice Off Date' and 'Year' for five different sections (Section 1 to Section 5). The data spans from 1875 to 1994. Overlaid on the spreadsheet is the 'Function Arguments' dialog box for the TTEST function. The dialog box shows 'Array1' as 'B2:B21' and 'Array2' as an empty array. The 'Tails' is set to 1, and the 'Type' is set to 2. The formula result is displayed as 0.172.

Section	Year	Ice Off Date
Section 1	1875	105
Section 1	1876	126
Section 1	1877	85
Section 1	1878	73
Section 1	1879	86
Section 1	1880	100
Section 1	1881	103
Section 1	1882	99
Section 1	1883	112
Section 1	1884	95
Section 1	1885	108
Section 1	1886	110
Section 1	1887	91
Section 1	1888	106
Section 1	1889	102
Section 1	1890	92
Section 1	1891	114
Section 1	1892	113
Section 1	1893	104
Section 1	1894	105
Section 2	1901	89
Section 2	1902	82
Section 2	1903	110
Section 2	1904	140
Section 2	1905	102
Section 2	1906	110
Section 2	1907	107
Section 2	1908	88
Section 2	1909	75
Section 2	1910	101
Section 2	1911	88
Section 2	1912	107
Section 2	1913	85
Section 2	1914	95
Section 2	1915	96
Section 2	1916	100
Section 2	1917	86
Section 2	1918	108
Section 2	1919	107
Section 2	1920	101
Section 2	1921	89
Section 2	1922	82
Section 2	1923	110
Section 2	1924	140
Section 2	1925	102
Section 2	1926	110
Section 2	1927	107
Section 2	1928	88
Section 2	1929	75
Section 2	1930	101
Section 2	1931	88
Section 2	1932	107
Section 2	1933	85
Section 2	1934	95
Section 2	1935	96
Section 2	1936	100
Section 2	1937	86
Section 2	1938	108
Section 2	1939	107
Section 2	1940	101
Section 2	1941	89
Section 2	1942	82
Section 2	1943	110
Section 2	1944	140
Section 2	1945	102
Section 2	1946	110
Section 2	1947	107
Section 2	1948	88
Section 2	1949	75
Section 2	1950	101
Section 2	1951	88
Section 2	1952	107
Section 2	1953	85
Section 2	1954	95
Section 2	1955	96
Section 2	1956	100
Section 2	1957	86
Section 2	1958	108
Section 2	1959	107
Section 2	1960	101
Section 2	1961	88
Section 2	1962	107
Section 2	1963	85
Section 2	1964	95
Section 2	1965	96
Section 2	1966	100
Section 2	1967	86
Section 2	1968	108
Section 2	1969	107
Section 2	1970	101
Section 2	1971	88
Section 2	1972	107
Section 2	1973	85
Section 2	1974	95
Section 2	1975	96
Section 2	1976	100
Section 2	1977	86
Section 2	1978	108
Section 2	1979	107
Section 2	1980	101
Section 2	1981	88
Section 2	1982	107
Section 2	1983	85
Section 2	1984	95
Section 2	1985	96
Section 2	1986	100
Section 2	1987	86
Section 2	1988	108
Section 2	1989	107
Section 2	1990	101
Section 2	1991	88
Section 2	1992	107
Section 2	1993	85
Section 2	1994	95

3. The first set of data or the first Array is the dates in section 1, B2:B21. You may either type in B2:B21 or you can highlight that section.
4. The second set of data is Section 2, D2:D21
5. This is a 1-tail T-test and a type 2 (equal variance) T-test. When you are done, the window should look like the window shown below, showing the T-test result of 0.172

Function Arguments

TTEST

Array1 B2:B21 = {105;126;85;73;86;100;103;99;1...}

Array2 D2:D21 = {101;107;68;102;85;123;80;103;106}

Tails 1 = 1

Type 2 = 2

= 0.172146584

Returns the probability associated with a Student's t-Test.

Type is the kind of t-test: paired = 1, two-sample equal variance (homoscedastic) = 2, two-sample unequal variance = 3.

Formula result = 0.172146584

[Help on this function](#) OK Cancel

6. A t-test of the type you have just performed, tells you the probability that the two data sets could be the result of chance. A t-test of 0.50 would indicate that there was a 50% chance that the differences between the two sets of data are the result of chance variations rather than the result of an actual warming process. However, a t-test result of 0.25 would indicate that there is a 25% probability that the data sets differ because of chance, and a 75% chance that the difference is due to a real effect. At that point the IPCC would call climate change being responsible for the difference between the two data sets as “Likely”. Below you will see the IPCC likelihood scale.

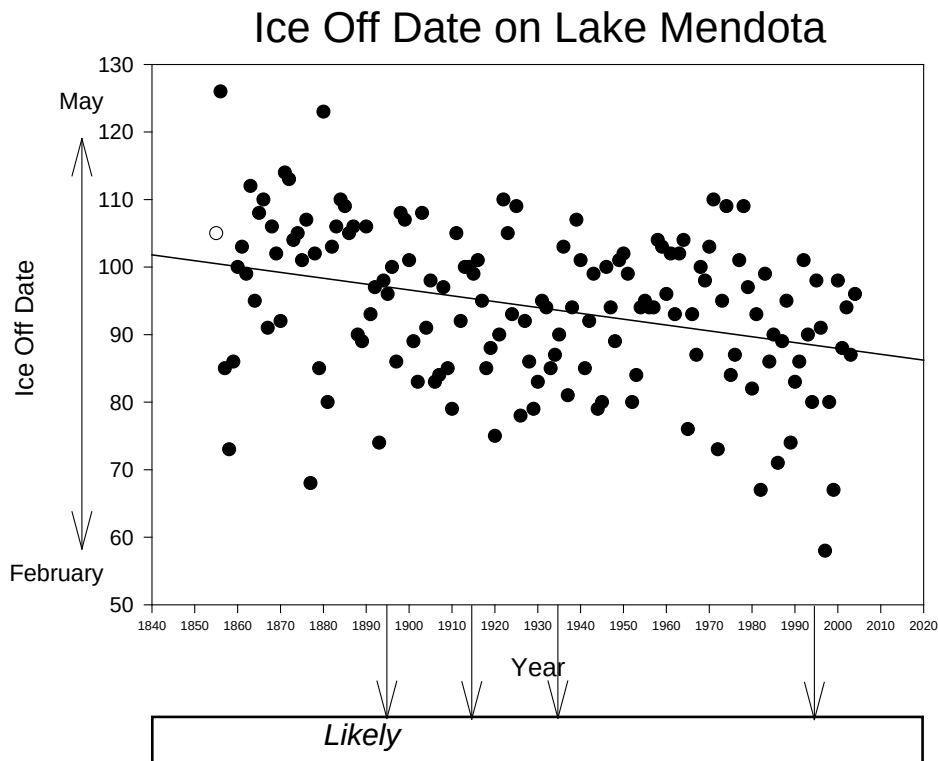
t-statistic (probability that data is a result of chance)	Probability of occurrence (from an effect rather than from chance)	IPCC Likelihood Scale
0.25	75%	Likely
0.20	80%	
0.15	85%	
0.10	90%	Very Likely
0.05	95%	Extremely Likely
0.025	97.5%	
0.02	98%	
0.01	99%	Virtually Certain
0.005	99.5%	
0.0025	99.75%	
0.001	99.9%	
0.0005	99.95%	
0.0001	99.99%	

- 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).
- Enter your results on a table like this:

Which comparison?	T Stat	Probability of Occurrence (not chance)	IPCC Likelihood Scale
1 vs. 2	0.172	83	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?