Iowa Science Teachers Journal

Volume 15 | Number 2 Article 19

1978

Don't Just Talk about the Weather: Investigate it Statistically

Bonnie H. Litwiller
University of Northern Iowa

David R. Duncan
University of Northern Iowa

Don't Just Talk About The Weather: Investigate It Statistically!

Bonnie H. Litwiller and David R. Duncan
Professors of Mathematics
University of Northern Iowa
Cedar Falls, Iowa 50613

Introduction

Scientists frequently wish to determine whether the means of two sets of data differ significantly. A statistical tool for investigating such a question is the "Student t-distribution" so named because its originator, W. S. Gosset, published his findings under the pseudonym "Student".

The problem considered in this paper will involve paired variates, that is, the case in which the observations in the first sample of data are paired naturally with those of the second. In this case the observations in the two samples of data can be written as shown in Table 1; the D column reflects the *differences* between paired observations while the D^2 column reflects the squares of these differences.

Table 1

	X	Y	D = (X - Y)	D^2
	x_1	y_1	$x_1 - y_1$	$(x_1 - y_1)^2$
	x_2	y_2	$x_2 - y_2$	$(x_2 - y_2)^2$
	x_3	y_3	$x_3 - y_3$	$(x_3 - y_3)^2$
	.	.	.	.
	.	.	.	.
	.	.	.	.
	x_n	y_n	$x_n - y_n$	$(x_n - y_n)^2$
Sum	Σx	Σy	ΣD	ΣD^2
Mean	$\overline{X}$	$\overline{Y}$	$D = (\overline{X} - \overline{Y})$	

Computations

To investigate whether the sample means $\overline{X}$ and $\overline{Y}$ differ significantly, compute the t-statistic using the following formula:

$$t = \frac{|D|}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{N}}{N(N-1)}}}$$

The statistical test then proceeds as follows:

1. Assume that the means of the populations from which the x's and y's are selected are the same; that is, assume that if the mean of all D's were known instead of only the mean for the observations in the sample, the mean of *all* the D's would be zero. This assumption is called the Null Hypothesis.

2. Compute the t-statistic as shown above.

3. Find the listed t-statistic on the Student t-distribution (Table 2). This is done by locating the degrees of freedom in the df column (for n pairs of observations, using (n – 1) degrees of freedom) and then following the row of numbers to the right until you intercept the desired level of significance. For example, at 4 df the table t-value is 2.776 for the .05 level of significance.

Table 2

Student's t-Distribution

A denotes the sum of the two tail areas for the values of t given below. ν denotes the number of degrees of freedom (df).

ν or df	A = 0.1	A = 0.05	A = 0.02	A = 0.01	A = 0.001
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.405
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.776	3.291

Alder and Roessler, p. 333

4. Compare the computed t-value with the t-value listed in the Table 2. If the computed t-value exceeds the table t-value, this indicates that the sample means $\overline{X}$ and $\overline{Y}$ differ by more than would be accounted for by chance alone if the Null Hypothesis were true. The Null Hypothesis is then rejected and the observed difference between $\overline{X}$ and $\overline{Y}$ is said to be significant. For instance, if the .05 significance level were used and if the Null Hypothesis were true, the probability that the computed t-value would exceed the table t-value by chance alone is only 5 chances in 100 or 5%. (If the .001 level of significance were used, the probability would be 0.1%.) Such a small probability would cause the rejection of the Null Hypothesis and the conclusion instead that the means of X and Y populations were in fact different and that the observed $\overline{X}$ and $\overline{Y}$ differ significantly.

5. If the computed t-value is less than the table t-value, there is not sufficient evidence to reject the Null Hypothesis; the observed $\overline{X}$ and $\overline{Y}$ are said *not* to differ significantly.

Weather Data

This statistical technique can be applied to weather data. In the weather data of 1974, the daily maximum temperatures in Waterloo seemed typically higher than those of Minneapolis. Were these differences statistically significant? Table 3 displays the Waterloo (W) daily maximum temperatures; the Minneapolis (M) daily maximum temperatures; difference W – M = D; D^2.

The alert reader will note $\sum D = \sum W - \sum M$ and $\overline{D} = \overline{W} - \overline{M}$. Why?

The t-statistic for significance of $\overline{D}$ for each month was then calculated using the data from Table 3 For instance for January.

$$t = \frac{|\overline{D}|}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{N}}{N(N-1)}}} = \frac{3.16}{\sqrt{\dfrac{1002 - \dfrac{(98)^2}{31}}{31(30)}}} = \frac{3.16}{\sqrt{.744}} = 3.66$$

At the .001 level of significance for 30 df (31 – 1), the table t-value is 3.646. Since 3.66 exceeds 3.646 it may be concluded that the difference between $\overline{W}$ and $\overline{M}$ for January 1974 was significant at the .001 level. Waterloo appeared to have a significantly different mean maximum temperature than Minneapolis. (A two-tailed test was used. This means that it was not presupposed which of $\overline{W}$ or $\overline{M}$ was the larger. The interested reader may consult a statistics text for the less common one-tailed test.)

On the other hand for April, the computed t is 3.12. Since the table t for 29 df is 2.756 at the .01 significance level and 3.659 at .001 significance level, it may be concluded the difference between $\overline{W}$ and $\overline{M}$ is significant at the .01 level but *not* at the .001 significance level.

Table 4 reflects the calculations of the data for January through December (1974).

Table 3

	January				February				March				April			
Date	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2
1	-3	-12	9	81	21	9	12	144	39	37	2	4	46	37	9	81
2	10	8	2	4	23	17	6	36	60	47	13	169	73	49	24	576
3	12	6	6	36	22	9	13	169	56	44	12	144	58	37	21	441
4	15	8	7	49	17	16	1	1	42	43	-1	1	41	41	0	0
5	12	8	4	16	22	19	3	9	53	53	0	0	51	46	5	25
6	6	4	2	4	20	18	2	4	66	54	12	144	61	51	10	100
7	0	0	0	0	20	14	6	36	52	40	12	144	54	42	12	144
8	8	3	5	25	15	9	6	36	49	36	13	169	46	42	4	16
9	3	1	2	4	20	21	-1	1	45	44	1	1	59	60	-1	1
10	10	7	3	9	29	29	0	0	45	47	-2	4	67	59	8	64
11	4	-4	8	64	44	39	5	25	36	45	-9	81	58	53	5	25
12	0	0	0	0	49	38	11	121	45	47	-2	4	55	60	-5	25
13	20	20	0	0	30	29	1	1	42	42	0	0	56	47	9	81
14	39	39	0	0	27	16	11	121	35	40	-5	25	46	53	-7	49
15	37	38	-1	1	29	31	-2	4	44	34	10	100	55	57	-2	4
16	43	41	2	4	39	30	9	81	31	30	1	1	60	59	1	1
17	34	33	1	1	47	38	9	81	33	33	0	0	67	69	-2	4
18	35	34	1	1	41	39	2	4	42	39	3	9	66	58	8	64
19	33	32	1	1	41	39	2	4	35	33	2	4	61	63	-2	4
20	34	34	0	0	43	37	6	36	28	29	-1	1	74	76	-2	4
21	33	34	-1	1	36	35	1	1	35	29	6	36	65	59	6	36
22	31	31	0	0	33	29	4	16	33	24	9	81	48	41	7	49
23	28	26	2	4	16	17	-1	1	18	12	6	36	52	56	-4	16
24	31	31	0	0	14	19	-5	25	20	20	0	0	62	64	-2	4
25	43	39	4	16	23	27	-4	16	49	43	6	36	77	78	-1	1
26	36	39	-3	9	38	35	3	9	41	33	8	64	81	83	-2	4
27	31	28	3	9	42	40	2	4	51	41	10	100	81	77	4	16
28	34	31	3	9	37	37	0	0	38	36	2	4	78	75	3	9
29	42	37	5	25					40	36	4	16	73	65	8	64
30	47	37	10	100					41	42	-1	1	72	64	8	64
31	32	9	23	529					46	37	9	81				
Sum	740	642	98	1002	838	736	102	986	1290	1170	120	1460	1843	1721	122	1972
Mean	23.9	20.7	3.16		29.9	26.3	3.64		41.6	37.7	3.87		61.4	57.4	4.07	

Table 3 (Cont.)

	May				June				July				August			
Date	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2
1	72	75	- 3	9	77	64	13	169	95	92	3	9	72	84	-12	144
2	75	77	- 2	4	74	74	0	0	97	92	5	25	74	68	6	36
3	62	60	2	4	85	87	- 2	4	88	85	3	9	62	57	5	25
4	64	56	8	64	83	83	0	0	82	83	- 1	1	76	79	- 3	9
5	70	62	8	64	82	83	- 1	1	85	90	- 5	25	78	83	- 5	25
6	57	58	- 1	1	75	68	7	49	86	91	- 5	25	81	80	1	1
7	47	53	- 6	36	66	65	1	1	93	97	- 4	16	81	80	1	1
8	45	53	- 8	64	70	78	- 8	64	96	101	- 5	25	82	79	3	9
9	48	44	4	16	77	68	9	81	95	92	3	9	81	81	0	0
10	58	54	4	16	72	68	4	16	86	85	1	1	82	73	9	81
11	67	63	4	16	71	69	2	4	76	79	- 3	9	75	74	1	1
12	58	59	- 1	1	75	75	0	0	89	89	0	0	86	78	8	64
13	57	45	12	144	79	78	1	1	92	99	- 7	49	77	80	- 3	9
14	60	53	7	49	87	80	7	49	86	86	0	0	82	83	- 1	1
15	59	59	0	0	69	65	4	16	81	84	- 3	9	82	82	0	0
16	62	72	-10	100	60	63	- 3	9	85	90	- 5	25	81	80	1	1
17	68	73	- 5	25	76	73	3	9	92	92	0	0	76	80	- 4	16
18	60	59	1	1	87	81	6	36	94	81	13	169	84	81	3	9
19	66	71	- 5	25	83	84	- 1	.1	88	96	- 8	64	87	88	- 1	1
20	81	84	- 3	9	93	82	11	121	88	87	1	1	89	90	- 1	1
21	82	79	3	9	88	81	7	49	90	87	3	9	88	82	6	36
22	78	69	9	81	73	74	- 1	1	84	89	- 5	25	80	68	12	144
23	69	57	12	144	72	73	- 1	1	85	86	- 1	1	74	73	1	1
24	63	58	5	25	73	76	- 3	9	85	87	- 2	4	79	80	- 1	1
25	65	65	0	0	78	78	0	0	82	89	- 7	49	80	86	- 6	36
26	67	71	- 4	16	77	80	- 3	9	85	86	- 1	1	91	81	10	100
27	67	71	- 4	16	81	82	- 1	1	79	86	- 7	49	73	75	- 2	4
28	77	75	2	4	84	88	- 4	16	87	84	3	9	73	70	3	9
29	79	77	2	4	88	88	0	0	77	76	1	1	74	71	3	9
30	68	69	- 1	1	87	82	5	25	79	78	1	1	77	73	4	16
31	71	70	1	1					82	81	1	1	71	68	3	9
Sum	2022	1991	31	949	2342	2290	52	742	2689	2720	-31	621	2448	2407	41	799
Mean	65.2	64.2	1.00		78.1	76.3	1.73		86.7	87.7	-1.00		79.0	77.6	1.32	

Table 3 (Cont.)

	September				October				November				December			
Date	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2	W	M	W-M=D	D^2
1	64	63	1	1	52	45	7	49	65	59	6	36	29	34	- 5	25
2	66	61	5	25	53	52	1	1	53	51	2	4	24	26	- 2	4
3	62	63	- 1	1	61	63	- 2	4	43	38	5	25	22	26	- 4	16
4	67	72	- 5	25	77	73	4	16	43	33	10	100	23	31	- 8	64
5	70	72	- 2	4	65	62	3	9	43	40	3	9	35	36	- 1	1
6	68	65	3	9	54	45	9	81	49	47	2	4	35	33	2	4
7	80	75	5	25	55	54	1	1	60	57	3	9	33	37	- 4	16
8	85	75	10	100	71	66	5	25	61	62	- 1	1	21	17	4	16
9	82	62	20	400	63	63	0	0	51	49	2	4	26	29	- 3	9
10	81	81	0	0	75	83	- 8	64	46	45	1	1	39	44	- 5	25
11	80	74	6	36	62	78	-16	256	50	47	3	9	39	43	- 4	16
12	62	55	7	49	58	57	1	1	38	38	0	0	40	40	0	0
13	61	57	4	16	51	54	- 3	9	32	30	2	4	35	36	- 1	1
14	74	77	- 3	9	52	51	1	1	29	30	- 1	1	34	33	1	1
15	74	71	3	9	56	54	2	4	39	36	3	9	34	34	0	0
16	76	77	- 1	1	63	62	1	1	48	31	17	289	32	33	- 1	1
17	83	73	10	100	67	59	8	64	56	53	3	9	22	19	3	9
18	80	82	- 2	4	61	56	5	25	44	41	3	9	28	25	3	9
19	75	66	9	81	62	52	10	100	53	45	8	64	26	26	0	0
20	71	66	5	25	50	47	3	9	41	35	6	36	25	24	1	1
21	60	48	12	144	61	65	- 4	16	40	36	4	16	23	24	- 1	1
22	55	50	5	25	71	62	9	81	56	52	4	16	37	34	3	9
23	69	62	7	49	67	62	5	25	50	43	7	49	35	31	4	16
24	73	73	0	0	74	71	3	9	37	33	4	16	22	21	1	1
25	77	71	6	36	63	57	6	36	32	26	6	36	23	21	2	4
26	87	85	2	4	66	64	2	4	39	32	7	49	36	30	6	36
27	66	70	- 4	16	72	70	2	4	41	30	11	121	35	32	3	9
28	60	54	6	36	70	66	4	16	25	27	- 2	4	39	38	1	1
29	72	66	6	36	65	61	4	16	32	32	0	0	33	32	1	1
30	54	50	4	16	63	60	3	9	30	32	- 2	4	36	37	- 1	1
31					66	64	2	4					34	32	2	4
Sum	2134	2016	118	1282	1946	1878	68	940	1326	1210	116	934	955	958	- 3	301
Mean	71.1	67.2	3.93		62.8	60.6	2.19		44.2	40.3	3.87		30.8	30.9	-0.10	

Table 4

Month	N Days in Month	ΣD	$\Sigma(D^2)$	$\overline{D}$	$\sqrt{\frac{\Sigma(D^2) - \frac{(\Sigma D)^2}{N}}{N(N - 1)}}$	Computed t	df (N - 1)	Level of Significance
January	31	98	1002	3.16	$\sqrt{.744}$ = .863	3.66	30	0.001
February	28	102	986	3.64	$\sqrt{.813}$ =..902	4.04	27	0.001
March	31	120	1460	3.87	$\sqrt{1.070}$ = 1.035	3.74	30	0.001
April	30	122	1972	4.07	$\sqrt{1.696}$ = 1.302	3.12	29	0.01
May	31	31	949	1.00	$\sqrt{.987}$ = .994	1.01	30	Not Sign.
June	30	52	742	1.73	$\sqrt{.749}$ = .866	2.00	29	0.1
July	31	-31	621	-1.00	$\sqrt{.634}$ = .796	1.26	30	Not Sign.
August	31	41	799	1.32	$\sqrt{.801}$ = .895	1.48	30	Not Sign.
September	30	118	1282	3.93	$\sqrt{.940}$ = .970	4.05	29	0.001
October	31	68	940	2.19	$\sqrt{.850}$ = .922	2.37	30	0.05
November	30	116	934	3.87	$\sqrt{.558}$ = .747	5.18	29	0.001
December	31	-3	301	-0.10	$\sqrt{.323}$ =. .569	0.18	30	Not Sign.

Conclusion

Table 4 reveals a persistent pattern of significance. For eight of the twelve months $\overline{W}$ and $\overline{M}$ differed significantly, the difference in every case favoring $\overline{W}$. The observation that Waterloo is warmer than Minneapolis appears to be supported.

The reader and his/her students may wish to:

1. Perform this same procedure for minimum temperatures for the same two cities. Weather information is available for a minimal charge from the National Climatic Center, Federal Building, Asheville, N.C. 28801. In addition, most weather bureaus will have recent official local reports.

2. Pick any two cities in the United States and secure weather data for them. Test the data for the significance of differences between maximums and minimums.

3. Make and test other weather related conjectures using the t-test. You may wish to consult a statistics text to learn the modification of the t-test if your data is not paired.

Reference

Alder, Henry L. and Roessler, Edward B., *Introduction to Probability and Statistics*, 5th edition, Freeman and Company, San Francisco, 1972, 363 pages.

* * *

A Teaching Calendar

"What can we do today?" is a question voiced by teachers at lesson plan time, by children on rainy days, and others with time on their hands. It used to be difficult to find new and interesting learning activities to fill this time, but with *Science Fun, Every Day in Every Way*, there are always answers at hand.

Did you know: On December 28, 1869, chewing gum was patented? On January 3, 1888, Marvin Chester Stone received a patent on artificial drinking straws made from paraffined manila paper? On January 12, 1879, milk was delivered to customers for the first time in glass bottles? These facts and related activities and project are what make *Science Fun* so useful.

The jumbo-sized wall calendar has listings and activities concentrating on science, with reading skills, mathematics, social studies, art and music highlighted. All activities have been classified in a subject index for quick reference. There is also an annotated list of games and additional references.

The calendar can be purchased from Acropolis Books Ltd., 2400 17th St. NW, Washington, DC 20009 for $9.50. Every day of the year is covered.

* * *

If you think education is expensive, try ignorance.
Derek Bok
President of Harvard University