Weather Data and Statistical Techniques

Noojin Walker

Department of Education
Austin Peay State University
Clarksville, Tennessee 37044

The weather can serve as an interesting focus for the study and application of statistics. Weather data is available, understandable, and commonly used. It can provide a wide range of opportunities to develop hypotheses, identify assumptions, select appropriate statistical techniques, and test hypotheses.

Most students have a familiarity with weather data such as temperature. They know that it is usually given as an integer, and a big number means hot and a smaller number means it is not. These data are easily obtained directly from a thermometer, from television reports, or from the daily newspaper where generally today's temperature is listed along with the temperature a year ago. The data can be collected day-by-day, or a previous month's data can be collected all at once.

Days can be classified subjectively as hot or cold, and the temperature can be used to confirm generalizations. "Gosh, it's unusually hot today. It's 93°." But is 93° unusual? "Hasn't this been a particularly cold March this year?" But is this year's mean low of 48 really colder than last year's 51°? Students learn that statistical techniques permit generalizations to be made with a certain degree of confidence that they are correct.

In the initial teaching of statistical techniques, the teacher and the book provide the data, pose the question, communicate the assumptions, suggest the hypothesis, and identify the statistical technique. This methodology of instruction is generally satisfactory for the first phase of learning; however, real life is not that patterned.

Curriculum and Evaluation Standards for School Mathematics (National Council of Teachers of Mathematics [NCTM], 1989) states that students should be encouraged to draw inferences from real-world situations. They should be given the opportunity to apply statistical tools to other academic subjects through the exploration of data. They need to have practice designing and implementing statistical experiments and interpreting and communicating the outcomes. This is particularly true for college-intending secondary students and is imperative for college students.

Weather data work well in what is referred to as the role of linkage "between the exactness of mathematics and the equivocal nature of a world dependent largely on the individual opinion" (NCTM, 1989, p. 167). As an example, students can be provided or can develop data depicting two years of daily high and low temperatures for the month of March as in Table 1 (Tennessee weather, 1989). They can be asked to structure a variety of statistical experiments to utilize various statistical techniques.

Students, or a group of students, could generalize that this year's March seems to be colder than last year's. They could decide that cold is defined by the daily low temperature, and the research hypothesis would be that this year's mean low temperature for March is significantly lower than last year's mean low temperature. They would assume that some temperatures for March are high and some are low but that the central tendency is to approach a normal distribution. They would elect to test this central tendency by utilizing as the statistical technique the test of the significance of difference between two means. They would test the null hypothesis at the level of confidence they set.

Still, another student or student group could decide that cold is determined by pairing the days of this year and last. If this year's low temperature is less than last year's, then it is colder. Their criteria for a colder year is based upon the number of days the low temperature this March is lower than the low temperature last March. An assumption is made that the working data would not correspond to a normal distribution. The experiment, therefore, would utilize a distribution-free statistic. Among the available choices, a nonparametric statistic such as the Sign Test could be selected. They would recognize that this test, based upon the binomial distribution, would not take into account the quantity of difference between the daily temperatures. A 25° difference would have no more weight than a 1° difference; an 81° day would be defined as cold when compared with an 82° day. They would option for a quick, simple, and less powerful statistic rather than a more powerful, albeit more complicated, statistic. They would be prepared to defend their decision.

Still, another student group could decide that they must define what cold means. They might define cold as when the low temperature is less than 40°. The research hypothesis would center around the question of how many days during this March was the daily low temperature less than 40° as compared against the number of days last year. They recognize the need for a distribution-free statistic and select the nonparametric statistic Chi-square as the technique. The Chi-square, like the Sign Test, is weak because it makes no distinction of how great the difference is between the paired temperatures. The Chi-square experiment is improved, however, by defining cold as 40° rather than it being any temperature that is less than the other year's.

Still another student group sees the statistical technique as

Table 1

Daily High and Low Temperatures (°F) for March

	1989				1988			
Day	High	Rank	Low	Rank	High	Rank	Low	Rank
1	46	3	43	14	66	15	49	24
2	59	10	30	2	60	5	57	27
3	71	17.5	43	14	69	18.5	59	28
4	72	19	54	19	75	27	61	29.5
5	67	15	55	21	74	26	44	17
6	58	8	38	9	86	31	52	26
7	50	4.5	31	3.5	47	1	40	7
8	53	6	37	7.5	54	2	41	10
9	56	7	43	14	60	5	42	13.5
10	45	2	33	5	66	15	42	13.5
11	44	1	37	7.5	66	15	49	7
12	50	4.5	23	1	69	18.5	48	22.5
13	59	10	31	3.5	65	13	47	21
14	62	12.5	35	6	69	18.5	45	18
15	59	10	40	10.5	71	23	42	13.5
16	62	12.5	49	17.5	55	3	43	16
17	71	17.5	40	10.5	61	8	33	1
18	80	23	55	21	73	24.5	39	4.5
19	80	23	57	23	62	10.5	51	25
20	64	14	47	16	60	5	37	3
21	69	16	42	12	70	21.5	34	2
22	76	21	49	17.5	80	28	48	22.5
23	86	25	55	21	82	29	61	29.5
24	88	27.5	65	28.5	83	30	66	31
25	88	27.5	62	25	61	8	41	10
26	87	26	65	28.5	62	10.5	46	19.5
27	90	30.5	64	27	73	24.5	41	10
28	90	30.5	68	31	61	8	46	19.5
29	89	29	67	30	63	12	40	7
30	80	23	63	25	70	21.5	39	4.5
31	73	20	58	24	69	18.5	42	13.5

a question of correlation. How well do this year's temperatures correlate with last year's? They define March as a transitional month. The weather begins cool and gets warmer as the month progresses. Under ideal circumstances, if the coolest day has a rank of 1, then the ranking of days according to their temperature would progress sequentially from 1 (coolest) to 31 (warmest). The test for the generalization, "The temperatures this month are about what we expect," could be to compare the actual rank of the days according to their temperature with the ideal ranking of the transitional month. Rank order correlation, rho, would be an appropriate technique. Still to do so, the students must develop definitions. They must decide whether to use the high temperatures or the low temperatures or a daily mean. In any case, the logic of the definition must be defended.

The point is that a month of temperature data for successive years provides a variety of opportunities for students to hypothesize, assume, define, defend, select, and test. The following are examples of generalizations that students can make from the daily temperature data in Table 1. Each generalization was tested by the statistical technique indicated.

Generalization: March 1989 was hotter than March 1988.

Hypothesis: No significant difference exists between the daily high temperature of the two years.

Considerations: The data are considered to be normally distributed around the central tendency.

Statistic: Mean difference = 10.67, standard deviation of difference = 14.89, standard error = 4.29, $t = 2.49$.

Conclusion: Reject the null hypothesis at .01 in a one-tail test.

Generalization: The rank order of March days by their increasing high temperatures correlates well with the ideal ranking of cool (1) to hot (31).

Hypothesis: No significant correlation exists between the ranking of March days with the ideal ranking.

Considerations: The test for the correlation of ordinal data requires a nonparametric statistic.

Statistic: Rank order correlation, rho = 0.75

Conclusion: Reject the null hypothesis at .01 in a one-tail test.

Generalization: March 1989 was colder than March 1988.

Hypothesis: No significant difference exists between the number of days in 1989 and 1988 having a low temperature less than 40°.

Considerations: The dichotomization of metric data necessitates the use of nonparametric statistics.

Statistic: Chi square = 3.84

Conclusion: Accept the null hypothesis at .05 in a one-tail test.

Generalization: It's 90° today; it will probably be 100° tomorrow.

Question: What was the probability of a 100° day in March 1989?

Considerations: The data approximates a normal distribution, and the probability of a specific value occurring can be determined.

Statistic: Standard normal probability with mean of 68.5; standard deviation = 14.46; 100° corresponds to 2.18 standard deviations.

Conclusion: The probability was 146 in 10,000.

To reiterate, a month's worth of daily temperature data provides a variety of opportunities for students to hypothesize and test hypotheses using familiar data. They learn to define criteria and to defend the logic of their definitions. Furthermore, they also practice the precise use of the language--learning to say exactly what they intend to say. The greatest benefit to the students, however, is that they do the thinking themselves rather than simply to respond to question 7 on page 138.

Reference List

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Tennessee weather. (1989, daily). *Leaf-Chronicle*, p. 2.