

MONTHLY WEATHER REVIEW

Editor, JAMES E. CASKEY, JR.

Volume 82
Number 1

JANUARY 1954

Closed March 15, 1954
Issued April 15, 1954

THE RATIONAL RELATIONSHIP BETWEEN HEATING DEGREE DAYS AND TEMPERATURE¹

H. C. S. THOM

Climatological Services Division, U. S. Weather Bureau, Washington 25, D. C.

[Manuscript received December 2, 1953]

ABSTRACT

The probability function of degree days below the base 65° F. is derived from the temperature probability function. Standard statistical analysis is applied to this function to obtain the relationship between mean degree days and mean temperature. This relationship is modified for use with available data and applied in the conversion of a monthly normal temperature for Detroit to the corresponding degree day normal.

INTRODUCTION

Almost from the time that heating degree days first came into use there has been a need for a rational relationship between temperature and degree-day statistics. The lack of such a relationship has always made it necessary to estimate degree-day means or normals from degree-day records which were often not available and tedious to compile. Temperature means, on the other hand, are already available for most stations and if not, are easy to compute from published data. Such a relationship makes degree-day statistics quickly available from any place with a temperature record. It also removes the difficulties associated with the lack of consistency between temperature and degree-day means which has been troublesome in the past. This has made it difficult to adjust degree-day means for a heterogeneous record. In the recent normals revision program of the Weather Bureau, for example, the usual arithmetical procedures could not be applied to obtain degree-day normals because of the numerous heterogeneities in the records at most stations. With a rational conversion formula available, properly adjusted temperature normals may be converted directly to degree-day normals with uniform consistency. More important than this use, perhaps, is the fact that the rational relationship is basic to the full development of the climatological analysis of degree-day data.

The study reported here is another phase [1] in the development of a general climatological analysis for degree days *below* a given base. With proper modification it may also be employed in the analysis of degree days *above* any base. The probability function of degree days derived here from the temperature distribution will form the basis for the later development of methods for obtaining degree-day probabilities.

THE TEMPERATURE FREQUENCY CURVE

In a previous paper [1] it was observed that the average temperatures of a particular day through a series of years have been found to have a normal probability or frequency function, or to be normally distributed. This probability function describes bell-shaped curves like those shown in figure 1 which are normal frequency curves on temperature scale t .

A normal probability function is known to be completely specified by its mean and standard deviation. The mean serves to locate the curve along the t axis while the standard deviation σ determines its scale, or how widely it is spread along the t axis. In figure 1 it is seen that both frequency curves are located by a mean temperature of 60° F. but have different scales or standard deviations. The curve with a standard deviation of 5.0 is spread out widely along the t axis while the curve with a standard deviation of 2.5 is more closely concentrated about the mean.

¹ Paper presented at 127th National Meeting of the American Meteorological Society, New York, N. Y., January 26, 1954.

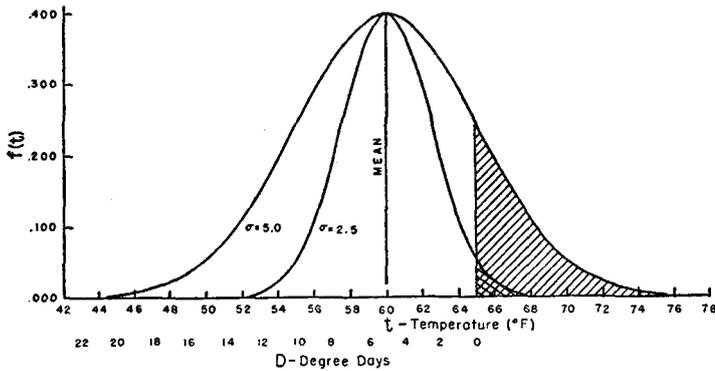


FIGURE 1.—Two examples of normal frequency curves, one for $\sigma=5.0$, the other for $\sigma=2.5$, both with mean temperatures of 60°F . Both the temperature scale t and its transformation by equation (1) to the degree-day scale D , are shown. As t is transformed to D , the distribution of t is transformed into the degree-day distribution. This is the unshaded portion under the temperature frequency curve distributed over the D scale together with an area of probability corresponding to the shaded portion concentrated at zero degree days. The entire distribution of degree days may be represented by the frequency curves shown in figure 2.

It is clear that as a result of these properties two changes may occur in the normal curve and hence in the distribution of temperature: (1) The mean may shift and move the curve to the left or right along the t axis, giving a location at a different value of t . (2) The scale or standard deviation may change causing the curve to spread out or become thinner. These changes are not statistically independent of each other but may be considered as separate component properties. An example of the first type of change is to move the curve $\sigma=2.5$ to the left two degrees of temperature, giving it a new mean of 58° but leaving the scale σ unchanged. The second type of change is represented in figure 1 by a change in scale from $\sigma=2.5$ to $\sigma=5.0$. This spreads the frequency curve without change in its location or mean. Also both types of change could occur together, giving a curve which is spread out as well as displaced along the t scale.

While the discussion of location and scale changes as climatic factors is a subject in itself, it will assist in our explanation of the degree-day distribution to have some understanding of climatic location and scale changes in the temperature distribution. The general principle observed over a wide range of climatic conditions is that the location of the temperature distribution increases as the scale decreases and conversely. This is in contrast to bounded elements such as precipitation where the location, as measured by the mean, varies directly as the scale. Since the location of the temperature distribution varies seasonally, as well as climatically, such variations are reflected in the seasonal march at a given station as well as from station to station for the same season.

The location and scale of the temperature distribution are best measured by the mean and standard deviation of the distribution. These parameters can therefore be related through the general principle. Although the variation of mean temperature with geographic position is not precise, there is, of course, a very marked tendency for it to decrease with increasing distance from the equator.

Since the mean and standard deviation vary inversely, the standard deviation increases with increasing distance from the equator. In general then, the mean decreases with latitude while the standard deviation increases with latitude. Similarly in seasonal variation the mean is higher in summer and lower in winter and hence the standard deviation is lower in summer and higher in winter.

Large bodies of water have a great effect on the relation between location and scale of the temperature distribution. The pronounced effects of decreasing the rate of change of mean temperature with latitude and the narrowing of the range between summer and winter are well known. The effect on the standard deviation is even more pronounced. As a consequence, standard deviations are stabilized over extended areas along seacoasts and through the seasons in such areas. For example, the standard deviation for January along the east coast of the United States is almost uniform from Maine to Florida while in the interior it is three times larger in Minnesota than in Louisiana. Seasonal variation in the standard deviation is also smaller along the coasts, some stations having nearly the same standard deviation the year around. This occurs particularly along the west coast where the effect is more pronounced because of the prevailing winds off the ocean.

THE DEGREE-DAY FREQUENCY CURVE

These location and scale changes in the temperature frequency distribution produce corresponding changes in the associated degree-day distribution. They may be illustrated by transforming temperature to degree days by the well-known relationship

$$D=65-t, D \geq 0 \quad (1)$$

where D is the degree-day value for a day and t is the day's average temperature in $^\circ\text{F}$. The inequality on the right is especially to be noted for it is an essential feature of the transformation which converts the t scale to the D scale of figure 1. As t is transformed to D , the distribution of t is transformed into the degree-day distribution. This is the unshaded portion under the temperature frequency curve distributed over the D scale together with an area or probability corresponding to the shaded portion concentrated at zero degree-days. Thus the probability of having degree days greater than zero on a particular day is equal to the unshaded portion below the temperature frequency curve and the probability of having zero degree days is the shaded portion. The manner in which these shaded and unshaded areas vary with the temperature distribution is clearly the key to the relation between temperature and degree-day statistics. Such variations may be interpreted in terms of the location and scale changes discussed above.

Since the degree-day base is fixed at 65°F ., all location and scale changes occur in relation to it. With fixed scale or standard deviation, shifts in the mean produce

important changes in the size of the shaded area. As the mean temperature increases, the temperature frequency curve moves toward the right and the shaded area of the curve is increased while the unshaded area is decreased. This produces an increase in the probability of zero degree days and both a decrease in probability of degree days and an increased concentration of the probability at the lower degree-day values. The overall effect is to decrease the mean degree days. For a decrease in mean temperature the shaded portion of the curve decreases while the unshaded portion increases. This produces a decrease in the probability of zero degree days and an increased concentration of probability at higher degree days with a consequent increase in degree days. As the temperature mean moves to low values on the left, the amount of shaded area becomes negligible and the degree-day mean approaches $65 - E(t)$ where $E(t)$ is the mean temperature. Thus, as has long been known, the degree-day mean increases as the temperature decreases and at low values is a function of the mean temperature alone. At higher values of mean temperature the shaded area becomes important and must be accounted for through use of both the mean and standard deviation since the size of the shaded area is a function of both parameters.

Variations in the degree-day mean produced by varying the temperature scale or standard deviation are not as easily depicted as those resulting from variation in the mean. With a fixed mean temperature, an increase in standard deviation increases the probability of zero degree days but also spreads the distribution to higher degree days. These changes have opposite effects on the degree-day mean so the effect of scale change is not a simple one and must be accounted for by an analytical relationship. Nevertheless, it is clear that changes in the temperature scale produce marked changes in the degree-day mean and hence must be accounted for in any relationship between degree days and temperature. As will be seen later, the scale or standard deviation is an important variable in the rational relationship.

THE PROBABILITY FUNCTION OF DEGREE DAYS

From the previous discussion it appears that the probability or frequency function of degree days consists of the portion of the temperature frequency curve below 65° and a probability concentrated at zero degree days equal to the probability of temperatures being above 65° . The former is the unshaded portion of the temperature frequency while the latter is equal to the shaded portion of the curve but concentrated at zero degree days. The unshaded portions of the frequency curves are truncated normal distributions which, when compounded with the probability densities at zero degree days, form mixed distributions which are the degree-day distributions. In sampling from such a distribution for a day on which zero degree days may occur, that day will have degree days greater than zero with a probability equal to the

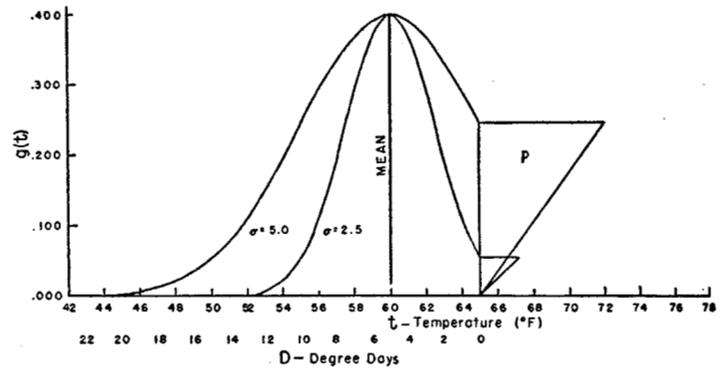


FIGURE 2.—Representation of the entire distribution of degree days, D , for two examples, corresponding to the two temperature frequency curves of figure 1 for which the standard deviations of temperature are 5.0 and 2.5, respectively, and mean temperature is 60° F. Note the area of probability, p , concentrated at zero degree days.

unshaded area of the frequency curve and zero degree days with a probability equal to the shaded portion of the curve. When degree-day values are greater than zero they will be further distributed according to the truncated probability function, the unshaded portion of the curve. They are not further distributed in the shaded portion of the curve, for here they always take the value zero.

The truncated normal distribution has been thoroughly investigated by several statisticians and most of the results we need have been reported in the literature (see [2, 3, 4, and 5]). There remains only to adapt the theory to cover the mixed distribution described above.

Let $F(t)$ be the normal distribution function of the average temperature for a day defined by

$$F(t) = \int_{-\infty}^t f(x) dx \tag{2}$$

where $f(x)$ is the normal probability function as shown in figure 1. Evidently $F(t)$ is the probability that an average temperature is less than t , and hence the probability that the average temperature is above the degree-day base is $p = 1 - F(65)$ and below the base is $q = F(65)$. Performing the transformation to degree days by equation (1), the distribution of degree days is

$$G(D|D \geq 0) = p + qF(65 - t|t \leq 65) \tag{3}$$

where G gives the probability of less than D degree days and F is the normal distribution truncated at 65° (c. f. [2]). It will be noted that $G(0) = p$ which is the probability of the average temperature being 65° or greater, and hence is the probability of zero degree days. When $D \geq 0$, G is equal to p plus the probability of temperature being between 65° and some assigned lower value.

The probability function for degree days is the derivative of (3) which is

$$g(D|D \geq 0) = qf(65 - t|t \leq 65). \tag{4}$$

This function is the equation of the unshaded portions of the curves in figure 1 referred to the D scale and is required in obtaining the mean value of D . The entire distribution may be represented by the frequency curves shown in figure 2.

THE RATIONAL RELATIONSHIP

The expected or mean value of degree days is defined in the usual manner by

$$E(D) = \int_0^{\infty} Dg(D)dD. \quad (5)$$

Applying this operation to the right hand side of equation (4) it is found that [2, 3]

$$E(D) = q[65 - E(t) + \lambda\sigma]. \quad (6)$$

Here $E(t)$ is the mean temperature, σ is the standard deviation [1], and $\lambda = f(65)/F(65)$. Tables of the reciprocal of this function have been prepared by Pearson [4].

Assuming that t is normally distributed, (6) is the exact relationship between mean temperature and mean degree days. Since $E(t)$ and σ completely define the normal distribution which in turn determines q and λ , the mean value of D is easily found when $E(t)$ and σ are known. Values of F and f are given as functions of the argument $(t - E(t))/\sigma$ in any table of the normal probability function. q and λ are evaluated at $t=65$ and for convenience we designate $(65 - E(t))/\sigma$ as h .

For $\sigma=5.0$ and $E(t)=60$ as shown in figure 1, it is seen that the base 65 is one standard deviation above the mean so, from tables of the normal distribution, $q=0.841$ and $\lambda=0.242/0.841=0.288$. Hence the degree-day mean for a day with $\sigma=5.0$ and $E(t)=60$ is

$$E(D) = .841[65 - 60 + (0.288)5] = 5.4.$$

APPLICATION OF THE RATIONAL RELATIONSHIP

The rational relationship applies to the means of daily degree days. However our interest is primarily in monthly means so the relationship will be adjusted to give these di-

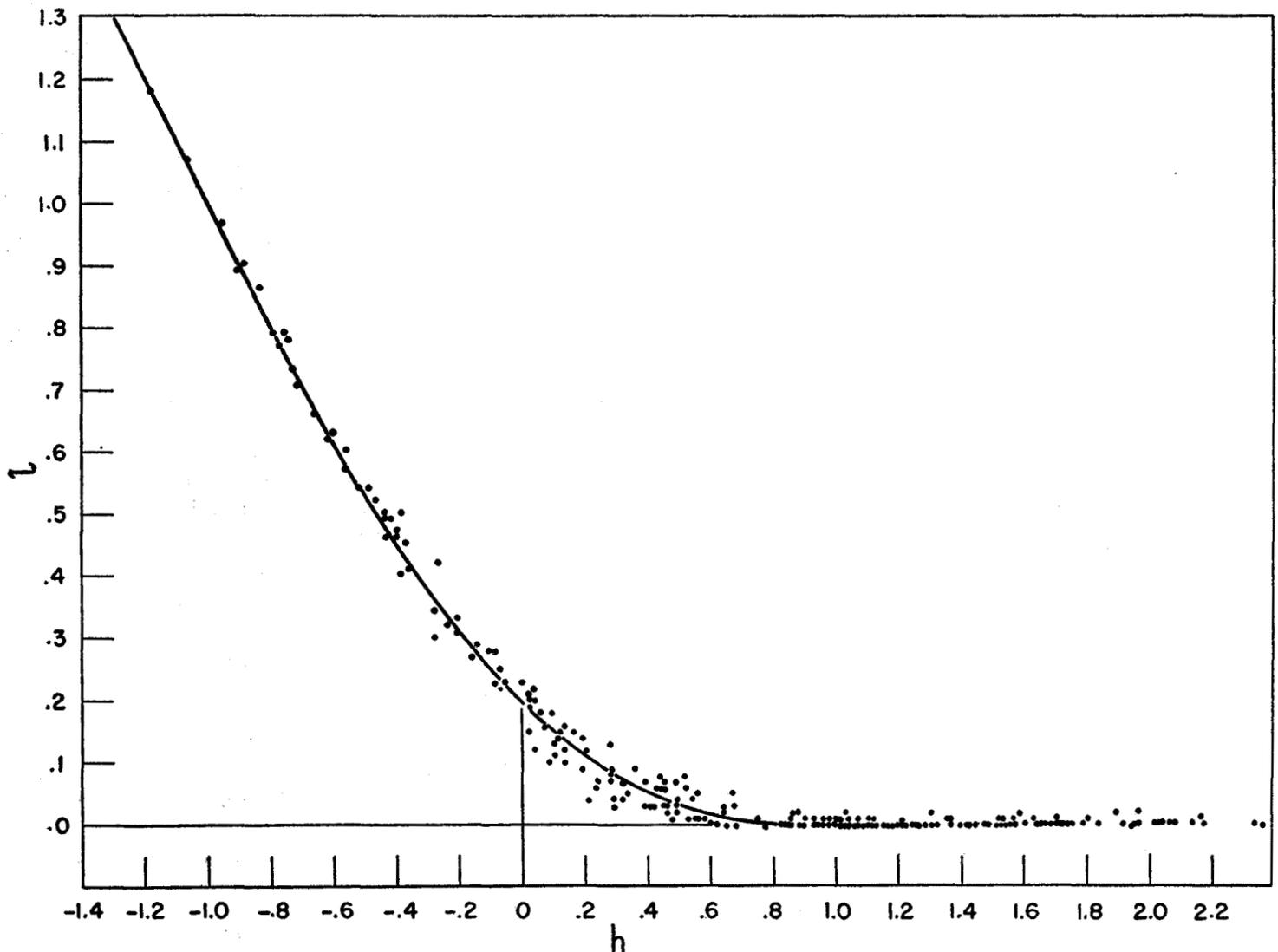


FIGURE 3.—The empirical relationship of l vs. h . The dots are observed values of l against h .

rectly. A simple way of doing this is to determine the relationship for a hypothetical average day of the month and multiply the resulting degree days by the number of days in the month. This average day is not a real day on which any particular average occurs, but a hypothetical day whose mean and standard deviation are such that when the conversion is made to degree days and the result multiplied by the number of days in the month the result is the mean degree days for the month.

In adjusting the relationship to obtain monthly statistics it was found convenient to use the standard deviation of monthly average temperature which is a function of the daily standard deviation and much easier to obtain. If σ is the standard deviation for the average day as above, σ_m the standard deviation of the monthly average, and r the mean correlation between all days for a month of N days, it may be shown [5] that

$$\sigma = \frac{\sqrt{N} \sigma_m}{\sqrt{1 + (N-1)r}}$$

Since the factor $(1 + Nr)$ is not known but does not seem to vary greatly from station to station, we let it be accounted for in the overall proportional adjustment to the rational relationship by assuming

$$\sigma = \sqrt{N} \sigma_m \tag{7}$$

Since σ for a single day is known only proportionally, q , which is a function of h , will also be known only proportionally. The approximation we need may be obtained by rearranging the rational relationship (6) in the form

$$\lambda - \frac{E(D)}{\sigma} \frac{(1-q)}{q} = \frac{E(D) - 65 + E(t)}{\sigma} \tag{8}$$

Substituting $\sqrt{N} \sigma_m$ for σ and l for the term on the left, we find

$$l = \frac{E(D) - 65 + E(t)}{\sqrt{N} \sigma_m} \tag{9}$$

Since all of the variables in (8) are functions of h , l will also be a function of h . Solving (9) for $NE(D)$, the mean monthly degree days, gives

$$NE(D) = N(65 - E(t) + l\sqrt{N} \sigma_m) \tag{10}$$

Next, l can be established as a function of h by plotting observed values of l against h . These values were computed from 30-year records at 30 stations representing all climatic conditions in the United States. The data which are for all of the 12 months are shown plotted on figure 3 together with the empirical l vs. h relationship. It is to be noted that the relationship is independent of climate and season and is only dependent on the parameters of the temperature frequency distribution. In this respect the l -function is general, like the λ -function, in that it is also dependent only on h . It is also similar in shape to the

TABLE 1.—The factors h and l , for use in computing degree days from equation (10)

h	l	h	l	h	l	h	l
-0.70	0.70	-0.32	0.39	0.05	0.17	0.42	0.05
-0.69	0.70	-0.31	0.38	0.06	0.17	0.43	0.05
-0.68	0.69	-0.30	0.38	0.07	0.16	0.44	0.04
-0.67	0.68	-0.29	0.37	0.08	0.16	0.45	0.04
-0.66	0.67	-0.28	0.36	0.09	0.15	0.46	0.04
-0.65	0.66	-0.27	0.36	0.10	0.15	0.47	0.04
-0.64	0.65	-0.26	0.35	0.11	0.14	0.48	0.04
-0.63	0.64	-0.25	0.34	0.12	0.14	0.49	0.03
-0.62	0.63	-0.24	0.34	0.13	0.13	0.50	0.03
-0.61	0.62	-0.23	0.33	0.14	0.13	0.51	0.03
-0.60	0.61	-0.22	0.32	0.15	0.13	0.52	0.03
-0.59	0.60	-0.21	0.32	0.16	0.12	0.53	0.03
-0.58	0.59	-0.20	0.31	0.17	0.12	0.54	0.03
-0.57	0.58	-0.19	0.30	0.18	0.11	0.55	0.03
-0.56	0.58	-0.18	0.30	0.19	0.11	0.56	0.02
-0.55	0.57	-0.17	0.29	0.20	0.11	0.57	0.02
-0.54	0.56	-0.16	0.29	0.21	0.10	0.58	0.02
-0.53	0.55	-0.15	0.28	0.22	0.10	0.59	0.02
-0.52	0.54	-0.14	0.27	0.23	0.10	0.60	0.02
-0.51	0.53	-0.13	0.27	0.24	0.09	0.61	0.02
-0.50	0.53	-0.12	0.26	0.25	0.09	0.62	0.02
-0.49	0.52	-0.11	0.25	0.26	0.09	0.63	0.02
-0.48	0.51	-0.10	0.25	0.27	0.08	0.64	0.02
-0.47	0.50	-0.09	0.24	0.28	0.08	0.65	0.01
-0.46	0.50	-0.08	0.24	0.29	0.08	0.66	0.01
-0.45	0.49	-0.07	0.23	0.30	0.07	0.67	0.01
-0.44	0.48	-0.06	0.23	0.31	0.07	0.68	0.01
-0.43	0.47	-0.05	0.22	0.32	0.07	0.69	0.01
-0.42	0.47	-0.04	0.22	0.33	0.07	0.70	0.01
-0.41	0.46	-0.03	0.21	0.34	0.06	0.71	0.01
-0.40	0.45	-0.02	0.20	0.35	0.06	0.72	0.01
-0.39	0.44	-0.01	0.20	0.36	0.06	0.73	0.01
-0.38	0.44	0.00	0.19	0.37	0.06	0.74	0.01
-0.37	0.43	0.01	0.19	0.38	0.06	0.75	0.01
-0.36	0.42	0.02	0.18	0.39	0.05	0.76	0.01
-0.35	0.41	0.03	0.18	0.40	0.05	0.77	0.01
-0.34	0.41	0.04	0.17	0.41	0.05	0.78	0.00
-0.33	0.40						

For $h \geq 0.78, l=0$
For $h \leq -0.70, l=-h$

λ -function and has analogous limiting properties, e. g., $l=-h$ for large values of $-h$, and $l=0$ for $h \geq 0.78$. Values read from figure 3 have been entered in table 1 for convenience in use.

In order to use (10) to compute normal monthly degree days, a set of manuscript charts has been prepared showing isolines of monthly standard deviations, s_m . Using the appropriate value of s_m and the normal value of the temperature, \bar{t} , as estimates of σ_m and $E(t)$, h may be readily calculated. Entering the table or graph with this value of h one finds the proper value of l . Substituting this together with \bar{t} and s_m in (10) and multiplying by N , the number of days in the month, gives the degree-day normal $N\bar{D}$ a statistical estimate of $NE(D)$.

As an example, for September at Detroit we find the normal temperature $\bar{t}=64.3$ and the standard deviation $s_m=2.7$. Then h is easily found to be $(65-64.3)/(5.48)(2.7)=0.047$. For this value of h table 1 gives $l=0.17$ and hence $\sqrt{N}ls_m=(5.48)(2.7)(0.17)=2.51$. Substituting in (10) gives

$$N\bar{D} = 30(65 - 64.3 + 2.51) = 96$$

This is Detroit's degree-day normal for September.

REFERENCES

1. H. C. S. Thom, "Seasonal Degree-day Statistics for the United States", *Monthly Weather Review*, vol. 80, No. 9, Sept. 1952, pp. 143-149.

2. Harold Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946, pp. 247-248.
3. R. A. Fisher, "Sampling Error of Estimated Deviates", etc., British Association, *Mathematical Tables*, vol. 1, 1931, p. xxxiii.
4. Karl Pearson (Editor), *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, London, 1931, pp. xxx and 11.
5. John F. Kenney, *Mathematics of Statistics*, Part II, D. Van Nostrand Co., New York, N. Y. 1939, p. 101.