Seasonal Degree-day Statistics for the United States[1]

BU-26-M | by H.C.S. Thom[2] | May, 1952

Ever since the advent of heating degree-days as a tool in the solution of heating design and operating problems the engineer has had only mean degree-day values readily available to him. A number of engineers had recognized the limitations of mean values, but little was done until the recent war to furnish more complete statistics. In their extensive operations in maintaining military installations, the Corps of Engineers soon recognized the unsuitability of the mean value for determining adequate fuel supplies for these installations. The main weakness of the mean in this application is its characteristic that it is exceeded as many times as not and hence the use of the mean value resulted in an equal probability of inadequate and plentiful annual fuel supplies. This clearly was too great a risk of inadequate supplies, so after some study of the problem, it was decided to use a statistic $x_{.75}$ which would be exceeded only 25% of the time. An arrangement was made with the U.S. Weather Bureau for producing the $x_{.75}$ and other degree-day statistics for about 650 military installations in the U.S. and Alaska. The Weather Bureau devised special methods under the supervision of the writer which made possible the rapid compilation of the required statistics. The use of these statistics resulted in considerable economy in fuels and transportation facilities and in more satisfactory operation of the heating facilities at the individual installations. The present results are an outgrowth of studies begun in connection with the development of $x_{.75}$ statistics. They are part of continuing studies which it

---

1. Presented at Annual Meetings, National District Heating Association, Skytop, Penna., June 4, 1952

2. Visiting Professor of Statistics, Biometrics Unit, Cornell University, and Chief Climatologist, U.S. Weather Bureau.

is hoped will eventually make possible the compilation of similar statistics for monthly data, for degree-days to any base, and the relation of these to mean temperature.

The mean value plays a double role in degree-day work with annual values: it serves as an estimate of the expected value, i.e. when multiplied by the number of years of a period it gives an estimate of total degree-days for that period and it serves to locate the frequency distribution of annual values along the degree-day scale. Actually both roles are intimately related to frequency distributions, for the fact that we can compute a mean value with any validity at all depends on the existence of a statistical population or frequency distribution. This implies that probabilities also exist and can be estimated. It is with the estimation of probabilities that we are principally concerned here.

The Frequency Distribution of Seasonal Degree-days.

The degree-day data used in this study were computed in the conventional manner using the formula

$$x = 65 - \bar{q}\ , \quad x > 0$$

where x is the degree-day value and $\bar{q}$ is the average temperature for a particular day. The x values were then summed to obtain the seasonal values.

It is a matter of observation that daily average temperatures for a particular day are approximately distributed in a normal frequency function. The corresponding daily degree-days will then be distributed approximately in a normal distribution which is truncated at 65° since by definition there will be no degree-days when the temperature is above 65°. The distribution of total seasonal degree-days will therefore be the combined distribution of some 200 to 300 truncated daily normal components. There is a theorem of statistical analysis which says that, under certain general restrictions met by these daily distributions, the sum of the daily values will approach a normal distribution as the number of days becomes large. Since 200 to 300 are large numbers in this respect, it is reasonable to expect that the distribution of seasonal degree-days could be closely approximated by the normal distribution.

In order to give more exact status to this hypothesis a statistical test for normality was applied. Geary and Pearson[3] have provided what they consider to be a powerful test of normality. This involves computing two statistics γ_1 and <u>a</u> which are measures of skewness and flatness and are given by

$$\gamma_1 = \frac{\Sigma(x - \bar{x})^3}{ns^3}$$

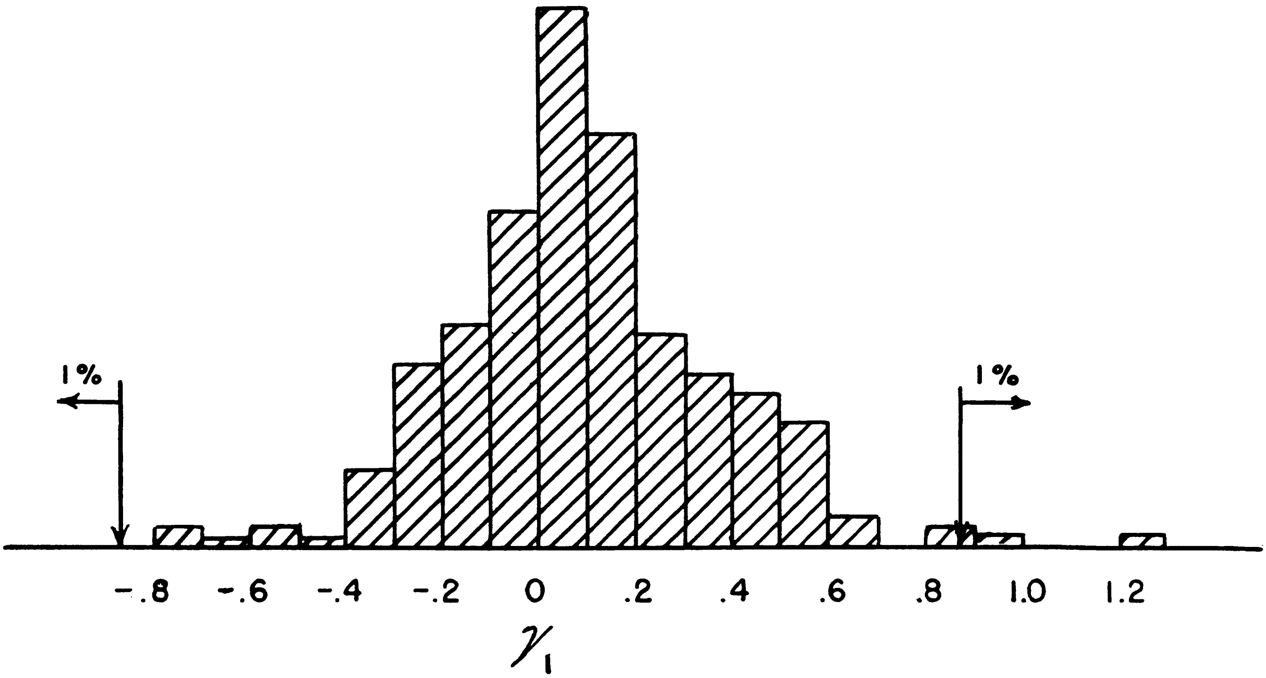
and

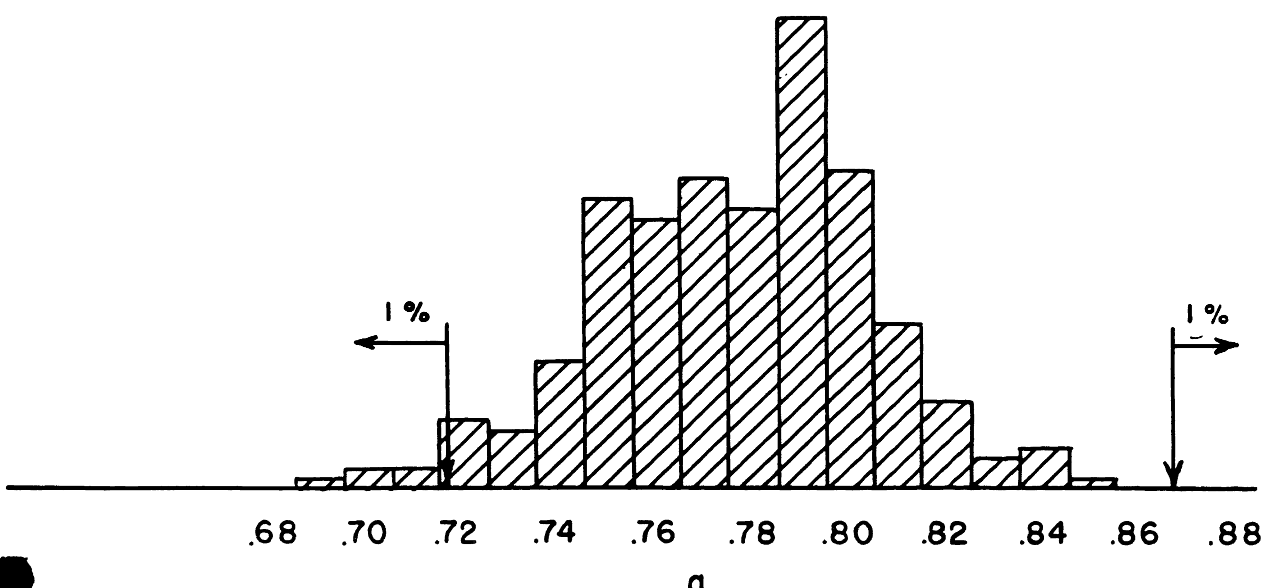
$$a = \frac{\Sigma|x - \bar{x}|}{ns}$$

where s is the standard deviation, n the length of record in years, and the summation extends over the years of record. These values were computed for the 266 weather stations used in this study and compared with Geary and Pearson's table to determine whether any were outside the limits allowed for normality. The results of this comparison are shown as histograms with the limits prescribed by the tables as vertical arrows in figure 1. These limits allow a total of 2% of the γ_1's and a's to lie outside the arrows. Since the lowest block is a frequency of one, it is seen that roughly the number of γ_1's and a's fall outside that would be expected if there were no departure from normality. Examination also showed that the larger departures were not related to climate. It was therefore concluded that the normal distribution could be successfully employed in fitting total seasonal degree-days and finally to estimate probabilities. Spot checks of actual frequency

3. Geary, R.C. and Pearson, E.S. "Tests of Normality," Biometrika Office, London (1938).

Figure 1





counts compared to normal estimates also verified this conclusion.

It is a well known principle of statistical analysis that the mean and standard deviation exhaust all of the information from a normal sample concerning the normal distribution in the population. It follows also through another principle that this applies to the estimation of probabilities. Hence if the distribution is normal any other technique for finding probabilities, such as plotting on probability paper, can be shown to waste part of the information available in the sample and to be therefore undesirable. All that is necessary then to completely define the statistics of seasonal degree-days for the United States are values or charts of the mean and standard deviation. These are shown in figures 2 and 3. Figure 3 also contains an abbreviated table of the normal probability distribution which facilitates the computation of 21 probability values. If other probabilities are required they may be readily computed using any one of a large number of normal probability tables.

Estimation of Probabilities and Quantiles for a Station.

Most normal tables[4] give

$$P = \int_0^t f(z)\,dz \quad \text{and} \quad f(t)$$

where $$f(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2\, z^2}$$

in terms of the argument

$$t = \frac{|x - \mu|}{\sigma} \cong \frac{x - \bar{x}}{s} \quad . \qquad (1)$$

μ and σ are estimated by $\bar{x}$ and s. The probability of being greater than a particular value t is $0.5 + P$ when $t < 0$ and $0.5 - P$ when $t > 0$ where P is the tabled result. For probabilities of being less than t the corresponding values of probability are $0.5 - P$ for $t < 0$ and $0.5 + P$ for $t > 0$.

---

4. "Handbook of Chemistry and Physics," Chemical Rubber Publishing Co.

Example 1. Required to find the probability that the season total of degree-days at Washington, D.C. will exceed 5000. From figure 2 we find that Washington has a degree-day mean of about 4500 and from figure 3 a standard deviation of about 390, hence

$$t = \frac{5000 - 4500}{390} = 1.28 .$$

The P value from a standard normal table is 0.40 and since $t > 0$ the probability is 0.5 - 0.40 or 0.10. The probability of being less than 5000 is immediately 1 - 0.10 or 0.90. Thus one tenth of the years in Washington will have degree-day totals of 5000 or greater and 9 out of 10 less than 5000.

Example 2. Required to find the seasonal degree-day total for Detroit which it would be unusual to exceed, i.e., which would be exceeded only with 0.05 probability or once in 20 years.

Referring to a standard normal table we find $t = 1.64$ for $0.5 - P = 0.05$. From figures 2 and 3 $\bar{x} = 7000$ and $s = 500$. Hence from equation (1)

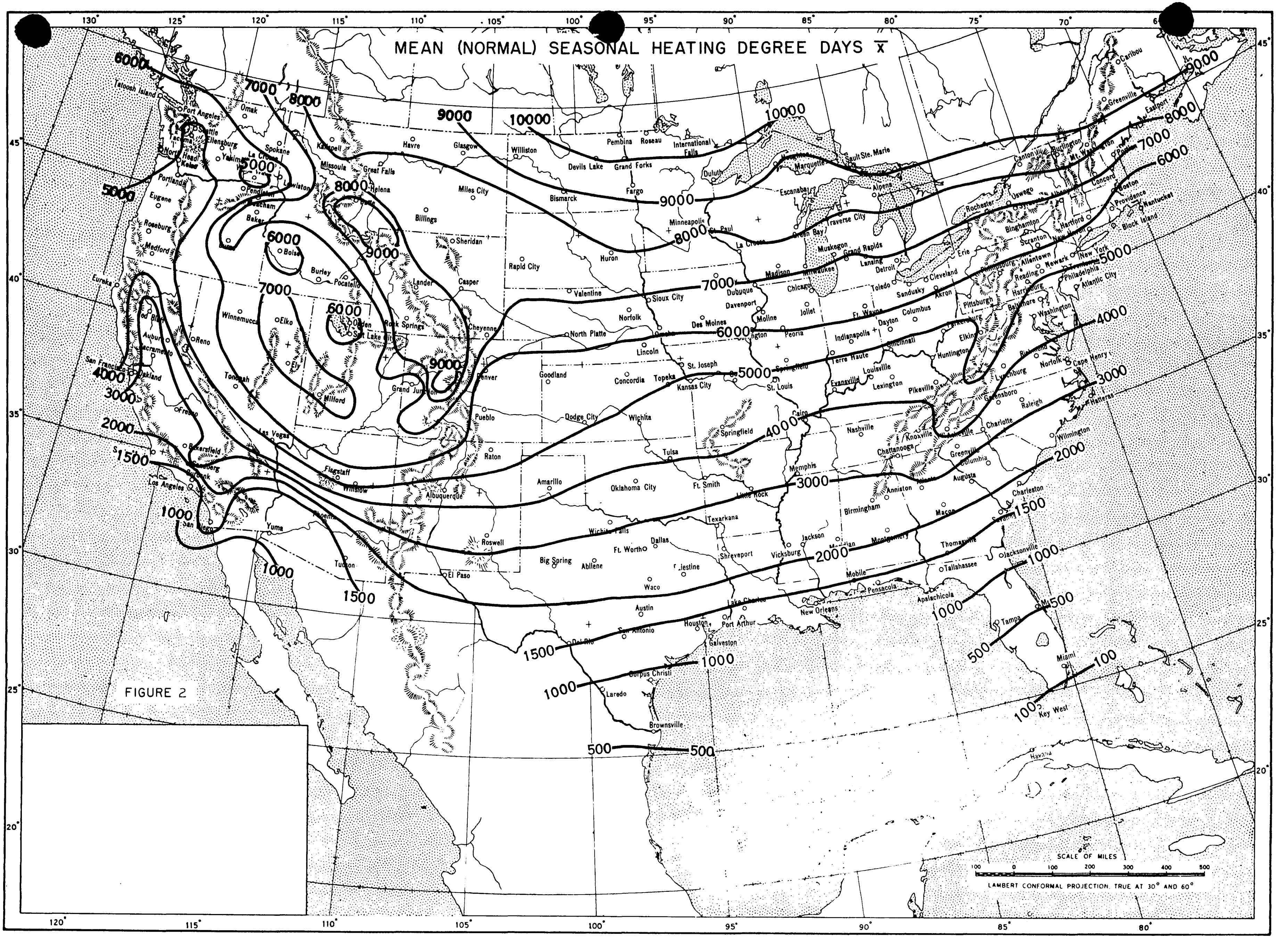
$$\begin{aligned} x &= ts + \bar{x} \\ &= 1.64 \times 500 + 7000 \\ &= 7820 \end{aligned}$$

Thus 7820 degrees for a season in Detroit would be an unusual value only exceeded once in 20 years on the average. This is called the $x_{.95}$ or the 0.95 quantile.

It should be noted that more accurate values of $\bar{x}$ may be used if available in conjunction with figure 3. More accurate values may also be ascertained for s at individual locations by the formula

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

although the drawing of isolines tends to enhance the accuracy of s since it varies slowly with latitude and is not as much influenced by local effects.



MEAN (NORMAL) SEASONAL HEATING DEGREE DAYS $\overline{X}$

FIGURE 2

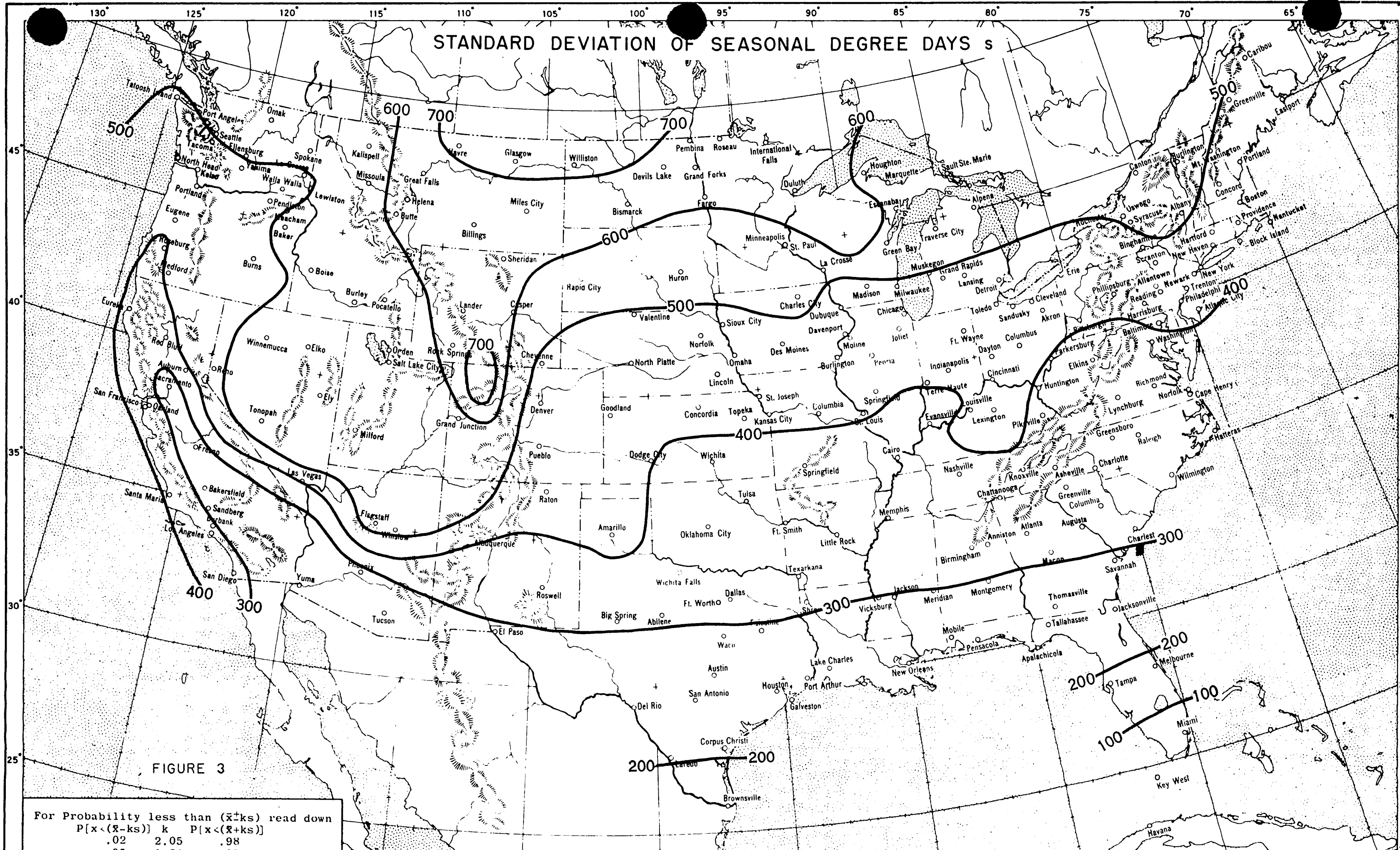


FIGURE 3

For Probability less than $(\bar{x} \pm ks)$ read down

$P[x<(\bar{x}-ks)]$	k	$P[x<(\bar{x}+ks)]$
.02	2.05	.98
.05	1.64	.95
.10	1.28	.90
.15	1.04	.85
.20	.84	.80
.25	.67	.75
.30	.52	.70
.35	.39	.65
.40	.25	.60
.45	.13	.55
.50	0	.50
$P[x>(\bar{x}+ks)]$	k	$P[x>(\bar{x}-ks)]$

For Probability greater than $(\bar{x} \mp ks)$ read up

Examples of the use of Figures 1 and 2

What are the heating-season degree-day totals at Kansas City for which (a) the probability of a season having a lesser value is 0.05 or 1 year in 20 and (b) the probability of a heating-season having a greater value is 0.10 or 1 in 10?

(a) $\bar{x}$ = mean value read from Figure 1 and s = standard deviation read from Figure 2. For Kansas City, $\bar{x}$ = 5,000, s = 410. Use column $P[x<(\bar{x}-ks)]$ and read down because 0.05 is in the first column and probability less than $(\bar{x}-ks)$ is wanted. k = 1.64 is found opposite 0.05. Substituting $\bar{x}-ks$ = 5,000 - 1.64x410 = 4330 degree-days which is the required value.

(b) For Kansas City again $\bar{x}$ = 5,000, s = 410. Use column $P[x>(\bar{x}+ks)]$ and read up since 0.10 is in the first column and probability greater than $(\bar{x}+ks)$ is wanted. k = 1.28 is found opposite 0.10 and $\bar{x}+ks$ = 5,000 + 1.28x410 = 5525 degree-days which is the required value.

Confidence Intervals for Probabilities and Quantiles[5]

After probabilities and quantiles are estimated by the procedures given above it is almost always desirable to have some measure of their reliability. The accepted statistical method of doing this is to calculate intervals which will enclose the true value with a prescribed probability or confidence. If these intervals are short our estimate of the true value is said to be accurate and, conversely, inaccurate if the intervals are long. Our opinion as to the length of the interval is a measure of the reliability of our estimates.

Fortunately both sample probabilities and quantiles are asymtotically normally distributed enabling us to use the normal distribution in determining approximate confidence limits. The inequality defining the limits for the 0.95 confidence interval of an estimated probability is

$$p - 2\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + 2\sqrt{\frac{p(1-p)}{n}}$$

The 0.95 confidence interval for an estimated quantile is

$$x_p - 2\sqrt{\frac{p(1-p)}{n}}\ \frac{s}{fp} \leq X_p \leq x_p + 2\sqrt{\frac{p(1-p)}{n}}\ \frac{s}{fp}$$

In these inequalities x_p is the quantile associated with the probability p, π is the true probability, X_p is the true quantile, and fp is the ordinate of the normal curve at x_p.

The confidence intervals for any example may now be readily calculated. For the first example

$$2\sqrt{\frac{p(1-p)}{n}} = 2\sqrt{\frac{0.1 \times 0.9}{51}} = .08$$

since 51 years of record were used for Washington, D.C. The 0.95 confidence interval is then

$$0.02 \leq \pi \leq 0.18$$

5. Kendall, M.G., "The Advanced Theory of Statistics," pp. 201 and 209, Vol. I (1947), Griffin and Co., London.

and we may say that the chances that the true probability is covered by this interval is 0.95.

Similarly for the $x_{.90}$ quantile

$$2\sqrt{\frac{p(1-p)}{n}}\ \frac{s}{fp} = .08 \times \frac{390}{0.1176} = 265$$

Hence the 0.95 confidence interval is

$$4735 \leq X_{.90} \leq 5265$$

and we may be confident that only in one chance in twenty will the true quantile X_p not be covered by this interval.

Confidence Intervals for the Mean or Normal.

Judging from the manner in which the mean or normal is often applied in degree-day work one cannot help but conclude that opinions as to its accuracy as an estimate of the true or population normal are somewhat exaggerated. Confidence limits established below will enable the engineer to form more exact opinions of the accuracy of the normals he uses.

Since the sample sizes or lengths of record ordinarily used in degree-day computations are of sufficient length for $\sqrt{n}\,(\bar{x} - \mu)/\,s$ to be normally distributed, confidence limits may be established using that distribution. The 0.95 confidence inequality for a mean or normal of degree-days is

$$\bar{x} - \frac{2s}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{2s}{\sqrt{n}}$$

where μ is the true normal and the other symbols are as previously. For Detroit we have

$$\frac{2s}{\sqrt{n}} = \frac{2 \times 500}{\sqrt{51}} = 140 \quad ;$$

hence the 0.95 confidence inequality is

$$6860 \leq \mu \leq 7140 \quad .$$

It is seen that the Detroit normal is no closer than 140 degree-days to the true value. Figure 3 may be used to judge the reliability of the seasonal normal for any station.

Summary.

Methods and charts are presented for estimating seasonal degree-day probabilities and quantiles for any location in the United States. Confidence limits for these are given enabling the engineer to judge the accuracy of these estimates and to apply them more effectively. A method for the determination of confidence limits for normals is also given. Examples for Detroit and Washington, D.C., are worked out using the methods presented.