FITTING A CHI-SQUARE CURVE TO AN OBSERVED FREQUENCY DISTRIBUTION

By W. T. Federer BU-14-M Jan. 17, 1951

Textbooks in statistics (for example, Johnson, Statistical Methods in Research; Love, Application of Statistical Methods to Agricultural Research; and Smith and Duncan, Sampling Statistics and Applications) are not clear as to the correct procedure for fitting a chi-square curve to an observed frequency distribution of sample variances. Therefore, the purpose of this note is to set out the procedure in detail with examples of 144 sample variances (Table 1) obtained by members of class in Advanced Statistical Methods, 213, and of 100 sample variances obtained by Johnson (Chapter III).

TABLE 1. 144 Sample Variances Obtained from Samples of Size 10 from the Population of Pig Gains, Table 3.1, Snedecor.

8.7	50.9	73.4	82.8	106.0	138.7
9.1	51.2	74.0	84.3	106.0	138.7
16.9	51.4	75.0	85.0	106.4	139.4
24.2	53.1	75.1	87.3	108.6	140.0
26.7	53.6	75.2	89.0	108.9	140.8
29.1	53.8	75.7	90.0	109.0	145.2
33.2	54.1	76.5	90.5	109.5	146.2
33.2	56.0	76.9	91.1	110.8	146.7
34.0	57.3	77.3	91.2	110.9	151.0
35.8	58.3	78.0	91.4	112.3	155.6
38.8	58.7	78.2	92.2	112.7	158.0
39.1	60.1	78.4	92.5	113.0	158.9
39.9	60.9	78.5	92.9	114.5	162.0
40.1	61.9	78.6	93.5	117.1	162.0
40.9	65.4	78.9	94.2	118.3	169.1
42.5	68.0	79.5	94.3	123.3	171.1
45.2	68.1	79.7	94.7	125.4	173.3
45.5	68.3	80.1	94.8	126.9	177.2
46.5	69.9	80.5	95.3	127.1	183.2
46.6	70.7	81.4	98.0	127.3	192.9
47.3	71.3	81.5	100.3	129.9	195.6
49.5	71.6	81.9	100.7	130.6	239.7
50.3	72.0	82.0	101.2	132.9	242.9
50.4	72.1	82.6	105.6	135.2	286.1
				Total:	13,400.8
				$\bar{x}$ =	93.061111

Ordinates for a chi-square curve with n degrees of freedom are obtained by substituting values of χ^2 in the formula

$$f(\chi^2) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} (\chi^2)^{\frac{n-2}{2}} e^{-\frac{\chi^2}{2}}$$

and evaluating the result.

Each of the 144 sample variances has 9 degrees of freedom associated with it. Therefore, it is desired to obtain ordinates for a chi-square curve with n equal to 9. The values of the ordinates are presented in Table 2 for selected values of chi-square.

TABLE 2. Computation of ordinates for χ^2 curve with 9 degrees of freedom.

Value of χ^2	$(\chi^2)^{\frac{7}{2}}$	$e^{-\frac{\chi^2}{2}}$	$.003799449(\chi^2)^{\frac{7}{2}} e^{-\frac{\chi^2}{2}}$ = ordinate *
0.0	0	1	0
1.0	1	0.606531	0.0023
2.0	11.3137	0.367879	0.0158
2.2	15.7935	0.329559	0.0198
2.4	21.4160	0.301194	0.0245
3.0	46.7654	0.223130	0.0396
4.0	128.000	0.135335	0.0658
5.0	279.508	0.082085	0.0872
6.0	529.090	0.049787	0.1001
7.0	907.493	0.030197	0.1041
8.0	1448.15	0.018136	0.0998
9.0	2187.00	0.011109	0.0923
10.0	3162.28	0.006738	0.0810
11.0	4414.43	0.004087	0.0685
12.0	5985.97	0.002479	0.0564
13.0	7921.40	0.001503	0.0452
14.0	10267.1	0.000912	0.0356
16.0	16384.0	0.000335	0.0209
18.0	24743.1	0.000123	0.0116
20.0	35777.1	0.000045	0.0061

$$* \quad \frac{1}{2^{\frac{9}{2}}\left(\frac{9}{2}-1\right)!} = \frac{1}{16\sqrt{2}\left(\frac{7}{2}\right)\left(\frac{5}{2}\right)\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}} = \frac{1}{105(\sqrt{2\pi})} = .003799449 .$$

Using class intervals of 10 the frequency distribution of the 144 sample variances is given in Table 3.

TABLE 3. Frequency distribution of 144 sample variances.

X_i = class center	f_i frequency	X_i = class center	f_i frequency	X_i = class center	f_i frequency
5	2	105	11	205	0
15	1	115	8	215	0
25	3	125	6	225	0
35	7	135	6	235	1
45	9	145	5	245	1
55	13	155	4	255	0
65	8	165	3	265	0
75	22	175	3	275	0
85	12	185	1	285	1
95	15	195	2	—	—
Total frequency Σf_i =					144

The χ^2 values multiplied by $\frac{\sigma^2}{n} = 10.340123$ and the corresponding ordinates for a frequency of 144 with a class interval of 10 are given in Table 4.

TABLE 4. χ^2 values and ordinates for the frequency distribution of Table 3.

Values of χ^2	$\frac{\sigma^2}{n}\chi^2$	139.26 x ordinate for χ^2 curve	Values of χ^2	$\frac{\sigma^2}{n}\chi^2$	139.26 x ordinate for χ^2 curve
0	0	0	8	82.72	13.90
1	10.34	0.32	9	93.06	12.85
2	20.68	2.20	10	103.40	11.28
2.2	22.75	2.76	11	113.74	9.54
2.4	24.82	3.41	12	124.08	7.85
3	31.02	5.51	13	134.42	6.29
4	41.36	9.16	14	144.76	4.96
5	51.70	12.14	16	165.44	2.91
6	62.04	13.94	18	186.12	1.62
7	72.38	14.50	20	206.80	0.85

For this example it will be assumed that the true variance is not given and that the mean of the variances, 93.061111, is a suitable estimate of the true variance, σ^2. The procedure usually followed in practice will be to take the mean of the sample variances equal to σ^2.

In Table 4 the factor 139.26 is multiplied by the ordinates of a χ^2 curve for n = 9 degrees of freedom. It is not necessary to have the χ^2 values correspond to class centers or class endpoints. The resulting curve through the new ordinates will then have the same area between the curve and the abscissa as is included in the histogram of the frequency distribution of variances. This factor is obtained as follows:

$$\frac{(\Sigma f_i)(\text{Class interval})(n = d.f.)}{\text{true variance}} = \frac{144(10)(9)}{93.061111} = 139.26 .$$

The abscissa for variances, s^2, corresponds to the abscissa for χ^2 multiplied by σ^2/n since $\chi^2 = \frac{ns^2}{\sigma^2}$ for sums of squares of normally distributed variates. Therefore, since an expansion factor is used for the χ^2 values along the abscissa, it is necessary to use the reciprocal of the factor for the values of the ordinate. The additional factor, frequency times class interval, is necessary in order to change the ordinate values from a curve with unit frequency to one with a total frequency, Σf_i, and a given class interval.

Mathematically the Jacobian of the transformation from χ^2 to s^2 ordinates involves the factor n/σ^2, thus

$$f(\chi^2)\, d\chi^2 = \frac{n}{\sigma^2} f(\chi^2)\, ds^2$$

or the ordinates for the s^2 curve are obtained as $\frac{n}{\sigma^2}$ times the ordinates of the χ^2 curve. The factor Σf_i (class interval) merely changes the unit area to a total area basis.

If the class interval were 20, the results given in Table 5 would be obtained.

TABLE 5. Frequency distribution and ordinates for curve with a class interval of 20.

X_i = class centers	f_i = frequency	Value of χ^2	$\frac{\sigma^2}{n}\chi^2$	*278.53 x ordinate for χ^2 curve
10	3	0	0	0
30	10	1	10.34	0.64
50	22	2	20.68	4.40
70	30	3	31.02	11.03
90	27	4	41.36	18.33
110	19	5	51.70	24.29
130	12	6	62.04	27.88
150	9	7	72.38	28.99
170	6	8	82.72	27.80
190	3	10	103.40	22.56
210	0	12	124.08	15.71
230	1	14	144.76	9.92
250	1	16	165.44	5.82
270	0	18	186.12	3.23
290	1	20	206.80	1.70
Σf_i =	144			

TABLE 5 (continued)

$$*278.53 = \frac{144(20)(9)}{93.061111} = \frac{(\Sigma f_i)(\text{class interval})(n)}{\text{true variance}} .$$

In Chapter III of "Statistical Methods in Research," Johnson presents a fitted chi-square curve, Figure 2, but gives no explanation for constructing the figure. The 100 sample variances (Johnson, Table 8) were obtained from the population of pig gains (Snedecor) from random samples of size 5. Johnson uses $\sigma^2 = 100$ obtained from the population of pig gains. This procedure is not always possible, since σ^2 will be unknown for the majority of populations from which observed frequency distributions are obtained. Ordinates for a chi-square curve with 4 degrees of freedom are computed in Table 6.

TABLE 6. Ordinates for a chi-square curve with 4 degrees of freedom.

Value of X^2	$e^{-X^2/2}$	$.25\ X^2\ e^{-X^2}$ = ordinate *
0	1	0
.1	.951229	.0238
.2	.904837	.0452
.3	.860708	.0646
.4	.818731	.0819
.5	.778801	.0974
.6	.740818	.1111
.7	.704688	.1233
.8	.670320	.1341
.9	.637628	.1435
1.0	.606531	.1516
1.2	.548812	.1646
1.4	.496585	.1738
1.6	.449329	.1797
1.8	.406570	.1830
2.0	.367879	.1839
2.2	.332871	.1831
2.4	.301194	.1807
2.8	.246597	.1726
3.2	.201897	.1615
4.0	.135335	.1353
5.0	.082085	.1026
6.0	.049787	.0747
7.0	.030197	.0528
8.0	.018316	.0366
9.0	.011109	.0250
10.0	.006738	.0168
11.0	.004087	.0112
12.0	.002479	.0074
16.0	.000335	.0013
20.0	.000045	.0002

TABLE 6 (continued)

$$* \frac{1}{2^{\frac{4}{2}} \Gamma\left(\frac{4}{2}\right)} = \frac{1}{4} = 0.25 \, .$$

Assuming that Johnson used a class interval of 20, the ordinates in the above table must be multiplied by $\frac{20(4)(100)}{100} = 80 = \frac{(\text{class interval})(n)(\Sigma f_i)}{\text{variance} = \sigma^2}$ Fitting a curve through these values gives a figure which agrees with Figure 2 in Johnson's book.

Also, the frequency distribution of the quantities, $\chi^2 = \frac{\text{sum of squares}}{\sigma^2}$, could be compared with the chi-square curve but quite often the sample variances s^2 will be available instead of the sums of squares. Therefore, the χ^2 values must be multiplied by σ^2/n in order to correspond to the values of the sample variances.

The comparison of an observed frequency distribution of sample variances with the theoretical chi-square distribution is exemplified in Table 10 of Johnson's book. He divides the observed and theoretical frequency distributions into 9 classes and compares them with a chi-square goodness of fit test. The resulting chi-square value has 9-1 = 8 degrees of freedom. If Johnson had used σ^2 equal to the mean of the 100 sample variances, 85.92, the resulting goodness of fit χ^2 value would <u>have had 7 degrees of freedom instead of 8</u> since σ^2 is estimated from the data.

In the event that Johnson had used the mean of the 100 sample means, 29.832 (Table 3), instead of the population mean, 30, and the average of the sample variances, 85.92, divided by 5 as $\sigma_{\bar{x}}^2$ instead of $\sigma_{\bar{x}}^2 = 100/5 = 20$, the resulting chi-square for comparing the observed with the expected frequencies, Table 4, would have had 10 - 3 = 7 degrees of freedom instead of 9 since the observed number must equal the expected number and two constants, m = 29.832 and $\sigma_{\bar{x}}^2 = 17.184$, were obtained from the data.

Another interesting point to note is that the expansion factor for the ordinates of the chi-square curve,

$$\frac{N(\text{class interval})}{\sigma^2/n} \quad ,$$

is of the same form as the expansion factor for fitting the normal curve to an observed frequency distribution,

$$\frac{N(\text{class interval})}{\text{sample standard deviation} = s} \quad .$$

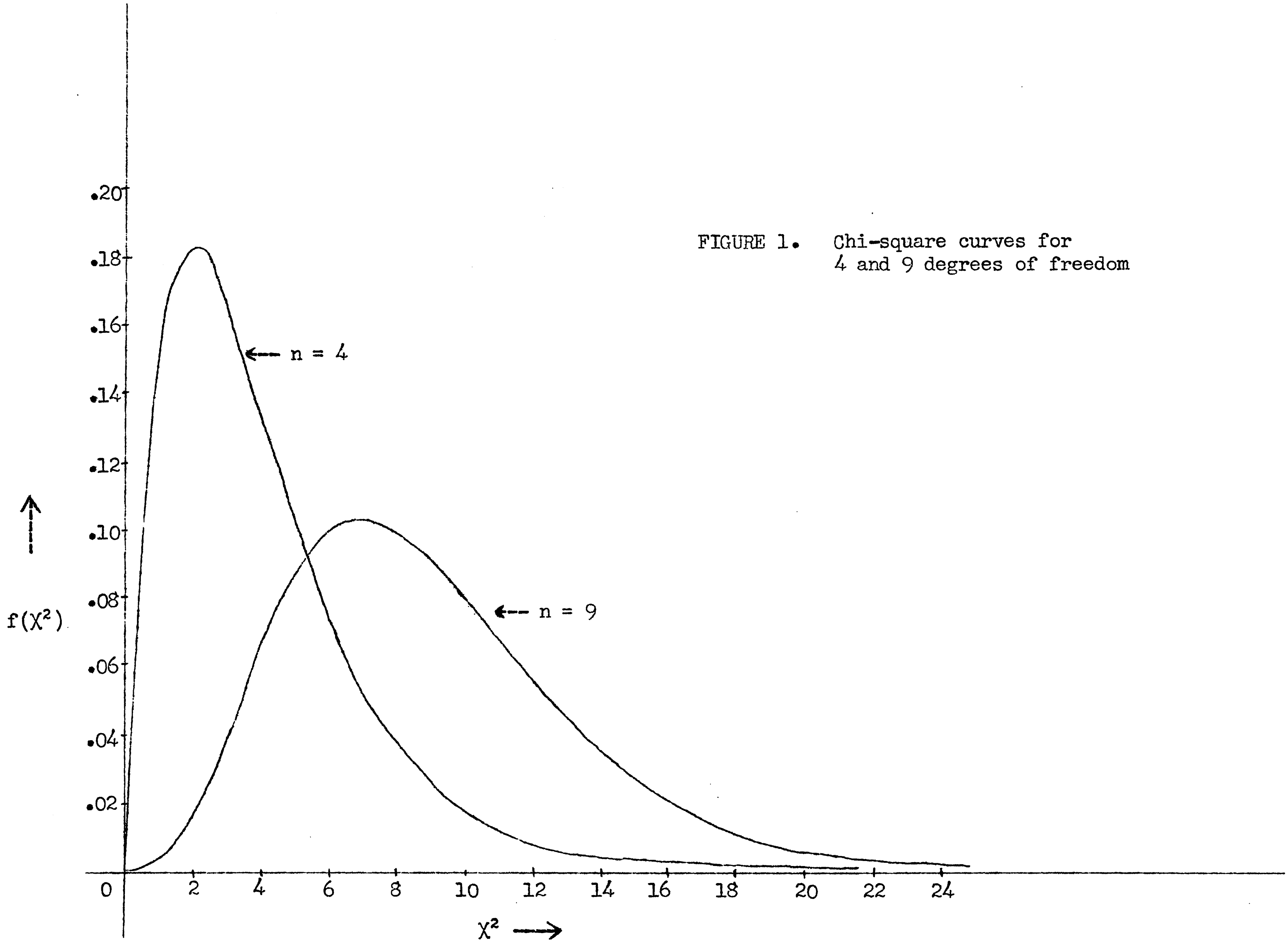

FIGURE 1. Chi-square curves for 4 and 9 degrees of freedom

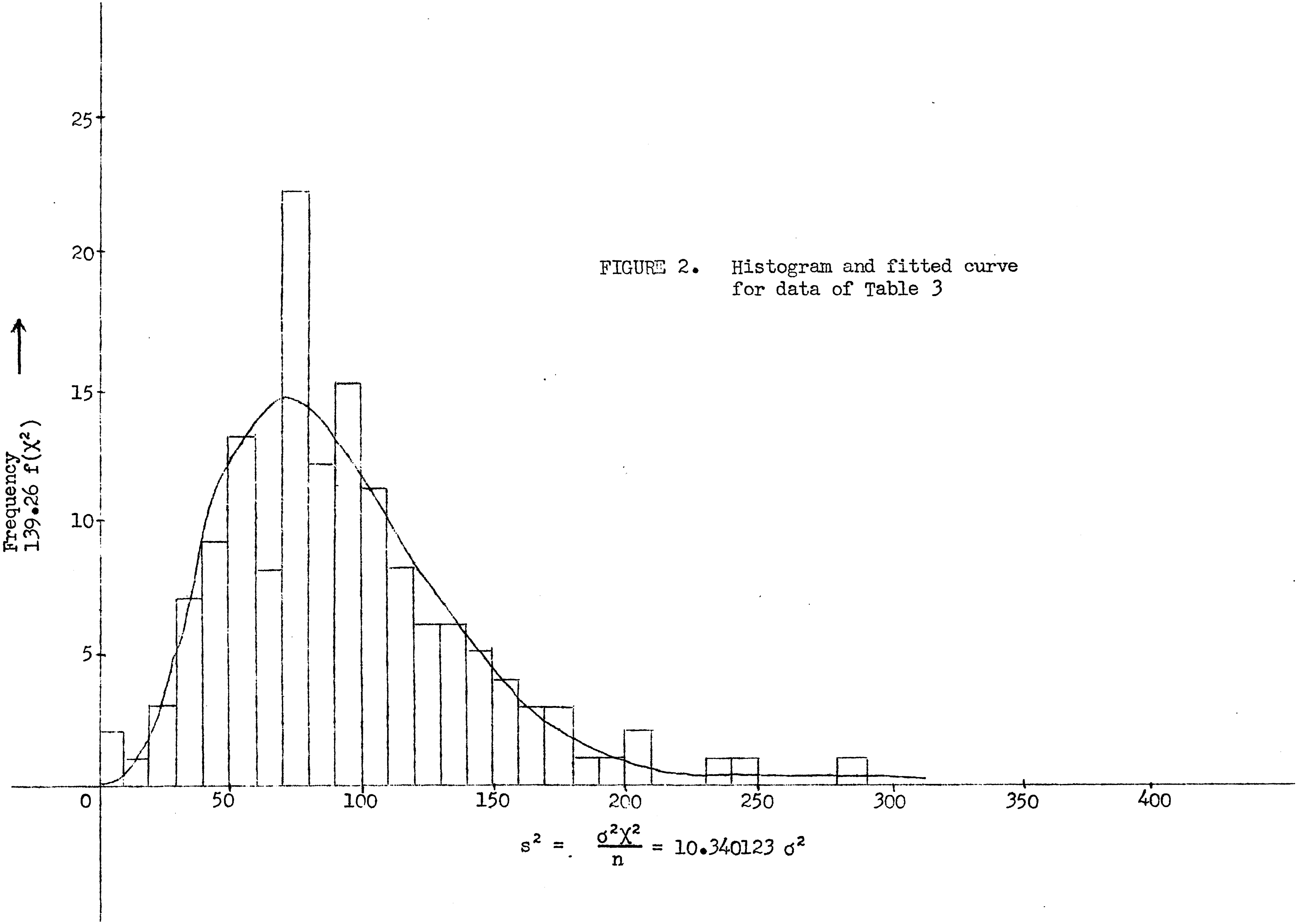

FIGURE 2. Histogram and fitted curve for data of Table 3

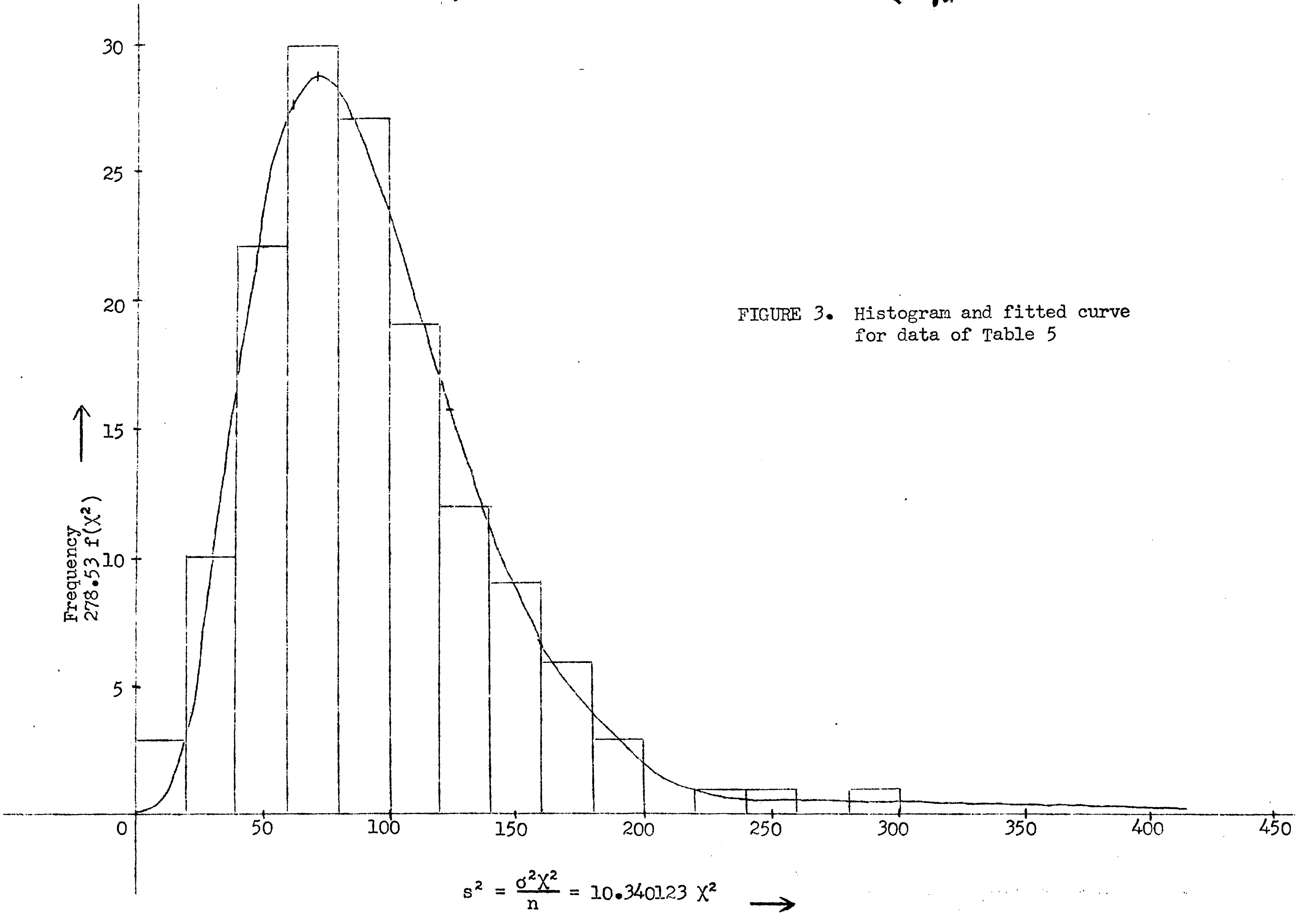

FIGURE 3. Histogram and fitted curve for data of Table 5