[Iowa Science Teachers Journal](https://scholarworks.uni.edu/istj)

Volume 14 | Number 2 Article 35

1977

Annual Baby Booms : A Chi-Square Investigation

David R. Duncan
University of Northern Iowa

Bonnie H. Litwiller
University of Northern Iowa

Follow this and additional works at: https://scholarworks.uni.edu/istj

Part of the Science and Mathematics Education Commons

Let us know how access to this document benefits you

Copyright © Copyright 1977 by the Iowa Academy of Science

Recommended Citation

Duncan, David R. and Litwiller, Bonnie H. (1977) "Annual Baby Booms : A Chi-Square Investigation," *Iowa Science Teachers Journal*: Vol. 14: No. 2, Article 35.
Available at: https://scholarworks.uni.edu/istj/vol14/iss2/35

This Article is brought to you for free and open access by the IAS Journals & Newsletters at UNI ScholarWorks. It has been accepted for inclusion in Iowa Science Teachers Journal by an authorized editor of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Offensive Materials Statement: Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.

ANNUAL BABY BOOMS : A CHI-SQUARE INVESTIGATION

David R. Duncan
Department of Mathematics
University of Northern Iowa
Cedar Falls, Iowa 50613

Bonnie H. Litwiller
Department of Mathematics
University of Northern Iowa
Cedar Falls, Iowa 50613

Introduction

Scientists frequently wish to determine whether an observed distribution of data is consistent with some theoretical distribution. For example, is a given set of data normally distributed, or is a given set of data "uniformily" distributed among various categories? A statistical tool for investigating these questions is the Chi-Square statistic.

Computation

To compute this statistic a set of data is partitioned into a finite number of categories (C) (each containing at least five observations) in some natural way. For instance, the data may be partitioned into categories according to their distances from the mean, their occurrence in months of the year, or numerous other ways. The number of data entries actually occurring in each of these categories is then recorded; these numbers are called the *observed frequencies* (0). Under the assumption of some theoretical distribution, the expected numbers of entries in each category are also computed; these numbers are called the expected frequencies (E). We then calculate $\frac{(O-E)^2}{E}$ for **each** category. The sum of these quotients for **all** the categories ($\sum \frac{(O-E)^2}{E}$) is called the **computed** Chi-Square statistic.

If the observed frequencies are very close to the expected frequencies in each category, then each of the fractions $\frac{(O-E)^2}{E}$ would be close to zero. The computed Chi-Square would then be quite small. If on the other hand, several of the observed frequencies differ markedly from the expected frequencies, the corresponding $\frac{(O-E)^2}{E}$ would be quite large as would be their sum -- the computed Chi-Square.

The statistical test then proceeds as follows:

1. Assume the observed distribution of the data is consistent with the theoretical distribution; call this assumption the null hypothesis.

2. Compute the Chi-Square as previously described.

3. Find the listed Chi-Square value in the Table 1. This is done by locating the degrees of freedom in the df column and then following the row of numbers to the right of the df value until you intercept the column indicating the level of significance desired in the interpretation of your results. For example, at 4 degrees of freedom, the Chi-Square value is 9.49 for a five percent level of significance (A = .05).

Table 1

Chi-Square Table [a]

ν or df	A = 0.30	A = 0.20	A = 0.10	A = 0.05	A = 0.02	A = 0.01	A = 0.001
1	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	18.42	20.46	23.54	26.30	29.63	32.00	39.25
17	19.51	21.62	24.77	27.59	31.00	33.41	40.79
18	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	24.94	27.30	30.81	33.92	37.66	30.29	48.27
23	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	33.53	36.25	40.26	43.77	47.96	50.89	59.70

[a] Alder and Roessler, p. 339.

4. Compare the computed Chi-Square value with the Chi-Square value listed in the table. If the computed Chi-Square exceeds the table Chi-Square entry, this indicates that the observed frequencies differ from the theoretical frequencies by more than could be accounted for by chance alone and the null hypothesis is rejected. For example, if the .05 level of significance were considered and if the null hypothesis were true, the probability that the computed Chi-Square would exceed the table Chi-Square by chance alone is only 5 chances in 100 or 5% (or 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 that the observed frequencies were sigificantly different from the expected frequencies. This would suggest that some factor is influencing the expected distribution of the frequencies.

5. If the computed Chi-Square is less than the table Chi-Square, there is not sufficient evidence to reject the null hypothesis.

This statistical technique can be applied to census data. It was noted in the natality section of recent census data that the numbers of live births per month appeared to vary. Were these variations significant?

U.S. Births

The U.S. census reports listing the numbers of live births per month for the years 1961 through 1970 were used. (Henceforth, "birth" means "live birth".) Within each year, comparisons were made for each month between the actual number of births and the "expected number" of births. The "expected number" of births was calculated by assuming the same number of births for each day of the year (a uniform distribution). For instance, there were 4,268,326 U.S. births registered for 1961. If the same number were born each day, there would be 4,268,326 $\div$ 365, or 11,694.04 births daily. Since there are 31 days in January, it would be expected that that there would be (to the nearest integer) 362,515 births during January (assuming uniform distribution of births); call this the *expected* number of births. There were, in fact, 363,286 registered births in January 1961; call this the actual or *observed* number of births. To test the significance of the differences between the actual and observed numbers of births, the Chi-Square test described earlier was used. Table 2 displays the observed and expected numbers of births for the months of 1961 (assuming uniform distribution) together with the necessary ratios to compute the Chi-Square statistic.

The null hypothesis was: the numbers of births per month were proportional to the numbers of days per month. The computed Chi-Square statistic used to test this hypothesis was 6,519.70. For (12 – 1) or 11 degrees of freedom and the .001 level of significance, the table Chi-Square statistic is 31.26. Consequently, the null hypothesis was rejected, that is, the numbers of births per month were *not* proportional to the numbers of days per month.

Table 2

U.S. Birth Data for 1961

Month	Observed Numbers of Births (O)	Expected Numbers of Births (E)	$\frac{(O-E)^2}{E}$
January	353,286	362,515	234.95
February	327,502	327,433	.01
March	360,322	362,515	13.27
April	335,120	350,821	702.70
May	342,404	362,515	1115.68
June	341,990	350,821	222.30
July	373,522	362,515	334.20
August	385,484	362,515	1455.32
September	377,628	350,821	2048.38
October	370,114	362,515	159.29
November	346,556	350,821	51.85
December	354,398	362,515	181.75
	4,268,326		6519.70

Tables 3 and 4 depict the situations for 1966 and 1970.

The computed Chi-Square for Table 3 is 4260.29. For 11 degrees of freedom and the .001 significance level, the table Chi-Square is 31.26. Thus the rejection of the null hypothesis is implied.

The computed Chi-Square for Table 4 is 7471.04. For 11 degrees of freedom and the .001 significance level, the table Chi-Square is 31.26. Thus the rejection of the null hypothesis is implied.

The data for '62, '63, '64, '65, '67, '68, '69 have been omitted; however, rejections of the null hypothesis also occur for these years.

An interesting pattern occurs for each month computed as to whether the actual number of births were greater or less than the expected number of births. Table 5 summarizes these data for the years 1961 through 1970. The

Table 3

U.S. Birth Data for 1966

Month	Observed Numbers of Births (O)	Expected Numbers of Births (E)	$\frac{(O - E)^2}{E}$
January	293,850	306,286	504.93
February	273,902	276,646	27.22
March	303,420	306,286	26.82
April	286,914	296,406	309.76
May	292,824	306,286	591.69
June	292,526	296,406	50.79
July	310,550	306,286	59.36
August	321,304	306,286	736.37
September	319,234	296,406	1758.12
October	312,942	306,286	144.64
November	296,458	296,406	.01
December	302,350	306,286	50.58
	3,606,274		4260.29

Table 4

U.S. Birth Data for 1970

Month	Observed Numbers of Births (O)	Expected Numbers of Births (E)	$\frac{(O - E)^2}{E}$
January	301,870	316,912	713.96
February	281,100	286,243	92.41
March	307,068	316,912	305.98
April	286,624	306,689	1312.74
May	297,648	316,912	1170.99
June	302,798	306,689	49.37
July	329,904	316,912	532.61
August	330,712	316,912	600.92
September	331,830	306,689	2060.95
October	323,764	316,912	148.15
November	309,012	306,689	17.60
December	329,056	316,912	465.36
	3,731,386		7471.04

entry G denotes that the actual number of births was greater than the expected number while the entry L indicates the actual number of births was less than the expected number.

Table 5

Results of Chi-Square Computations 1961-1970.

Months	'61	'62	'63	'64	'65	'66	'67	'68	'69	'70
January	L	L	L	L	L	L	L	L	L	L
February	G	L	L	L	L	L	L	L	L	L
March	L	L	L	L	L	L	L	L	L	L
April	L	L	L	L	L	L	L	L	L	L
May	L	L	L	L	L	L	L	L	L	L
June	L	L	L	G	L	L	G	G	L	L
July	G	G	G	G	G	G	G	G	G	G
August	G	G	G	G	G	G	G	G	G	G
September	G	G	G	G	G	G	G	G	G	G
October	G	G	G	G	G	G	G	G	G	G
November	L	L	L	L	L	G	L	G	G	G
December	L	L	L	L	L	L	L	G	G	G

Conclusion

A consistent pattern appears to emerge; there appears to be a "baby boom" in the late summer and early fall. The pattern for June appears to be erratic. In the last few years of the decade November and December became popular baby months (perhaps to save on income tax?).

The reader and his/her students are encouraged to:

1. Supply reasons explaining the persistent trends cited here.
2. Make and test other conjectures concerning census data.
3. Consult statistics books to learn of additional applications of Chi-Square such as tests of normality and contingency tables.

Selected Bibliography

Adler, H. L. and Roessler, E. B. *Introduction to Probability and Statistics.* W. H. Freeman, Fifth edition, 1972.

Freund, J. A. *Modern Elementary Statistics.* Prentice Hall, Third edition, 1967.