UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

AN ESTIMATING FORMULA AND THE RATIO METHOD OF ESTIMATING

Prepared by
F. H. Harper and Violet M. Feild

Washington, D. C.
October 1939

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

An Estimating Formula and the Ratio Method of Estimating

An Office Report
prepared by
F. H. Harper and Violet M. Feild

Basis of Estimates

The making of estimates on the characteristics of an entire statistical universe, or population, from sample data has already been described briefly in the chapter on Sampling. Because of the law of statistical regularity we know that frequently dependable estimates on the characteristics of an entire universe can be made on the basis of characteristics of a very small part of the universe if this small part is properly selected. If this were not possible, many of the reports that are now available on business conditions, costs of production, living costs, housing, and so on would not have been available because of the cost, effort, and time that would have been necessary to include the entire universe in the sample.

Many government and private agencies use sample data in making estimates, and these estimates are generally acceptable. Those who prepare the estimates and those who use them know that generally it is not necessary to include every part of the universe in the sample in order to obtain a basis for generalizing on the characteristics of the universe in its entirety, and they know also that in many instances the importance of

estimates would be minimized if it were necessary to withhold them until a 100 percent sample could be procured. Furthermore, as indicated, the time that would be required to procure a 100 percent sample frequently would prohibit the making of timely estimates on conditions in which there is a current interest.

This report is concerned with estimates that are somewhat different in nature from those that are made from sample data, by raising it to the universe, and the nature of such estimates has been indicated in the discussions on correlation and regression. For example, the fundamental principle underlying the calculation of numerical expressions of cause and effect relationships is that from the determination of the extent to which independent and dependent variables are associated, or have been associated in the past, logical deductions sometimes can be made as regards future relationships. When we have measured the degree of apparent cause and effect relationships between certain variables it is often possible to estimate or predict future occurrences. This is a principle underlying many correlation analyses, for it is only on the basis of past occurrences that we can arrive at conclusions as to what is most likely to happen in the future.

Simple regression lines enable us to show the relationship between a series of independent variables and a series of dependent variables, thus making it possible to make some deduction as regards most probable values of a dependent variable that are likely to be associated with given values of the independent variable. In like manner, the multiple regression equation provides a means of estimating the most probable values of a dependent variable that are likely to be associated with given values of 2 or more independents. By simple regression it might be possible, for example, to

estimate with some degree of dependability the price of cotton that is likely to be associated with a given quantity of production, and fairly dependable estimates of the yield of wheat per acre might sometimes be made by application of multiple regression methods involving the measurement of causal effects of 2 or more weather factors.

The method of estimating from regression coefficients has already been explained elsewhere, but some additional explanation may further clarify the necessary calculations. An assumption is, of course, that any estimating formula that makes it possible to determine most probable values of the dependent variable that would logically have been expected to be associated with given independent variables can be used with some degree of safety in estimating probable future occurrences, with the necessary changes being made in the calculations to obtain the desired measures of correlation and regression. For example, if our estimates of the acreage of cotton that would logically have been expected to be harvested annually during the period 1900-20 are close approximations of the acreages actually harvested the same formula, with the proper substitutions, can probably be used in estimating for future years.

If we correlate prices of 1899 to 1919 with acreages harvested from plantings in 1900-20, calculate the coefficient of regression, and then make estimates of acreages harvested that approximate the actual acreages harvested, then it might be possible to use the measures of relationship between prices of 1900-20 and harvested acreages of 1901-21 in estimating the acreage likely to be harvested from plantings of 1922, and so on for other years.

The Estimating Formula

Estimates frequently might be improved by taking into account the fact that the dependent variable may not be expected to coincide with the ordinate of the line of calculated trend of relationship. To illustrate the

application of the estimating formula and the results obtained thereby, tables 1-5 are included.

Table 1 shows index numbers of farm prices, seasonal average prices received for cotton by farmers, and deflated seasonal average prices. The actual prices in column 2 were divided by the corresponding index numbers in column 1 to obtain the deflated prices in column 3.

Table 2 shows the calculations made to determine the coefficient of correlation between production of cotton and seasonal average price, and table 3 shows the calculations made to determine the coefficient of correlation between seasonal average price and acreage harvested from the subsequent year's plantings. As will be observed, the coefficients of correlation are determined by the method of percentage change of first differences.

Table 1. - Index numbers of farm prices and seasonal average prices per pound received by farmers for cotton, 1910-27

Year	1	2	3
	Index number of farm prices 1/	Seasonal average price in cents per pound received by farmers for cotton lint 2/	Deflated seasonal average price in cents per pound received by farmers for cotton lint 3/
1910	102	13.95	13.68
1911	95	9.60	10.11
1912	100	11.49	11.49
1913	101	12.51	12.39
1914	101	7.36	7.29
1915	98	11.22	11.45
1916	118	17.33	14.69
1917	175	27.12	15.50
1918	202	28.92	14.32
1919	213	35.41	16.62
1920	211	15.92	7.55
1921	125	17.01	13.61
1922	132	22.87	17.33
1923	142	28.69	20.20
1924	143	22.91	16.02
1925	156	19.59	12.56
1926	145	12.47	8.60
1927	139	20.19	14.53

Table 1 - (continued)

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 566, page 444.
2/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95.
3/ Calculated by dividing seasonal average price by the corresponding index number of farm prices.

Table 2. - Production of cotton in the United States and deflated seasonal average price per pound received by farmers, 1910-27

Year	1	2	3	4	5	6
	Production					
	Thousands of bales 1/	First difference, or deviation from preceding year, in thousands of bales	Percentage change (mean = 3.0)	Deviation of percentage change from mean	Square of deviation	Multiple of standard deviation (S. D. = 20.87)
	-	-	-	-	-	x
1910	11,609	---	---	---	---	---
1911	15,694	4,085	35.2	32.2	1036.84	1.543
1912	13,703	- 1,991	- 12.7	- 15.7	246.49	- .752
1913	14,153	450	3.3	.3	.09	.014
1914	16,112	1,959	13.8	10.8	116.64	.517
1915	11,172	- 4,940	- 30.7	- 33.7	1135.69	-1.615
1916	11,448	276	2.5	- .5	.25	- .024
1917	11,284	- 164	- 1.4	- 4.4	19.36	- .211
1918	12,018	734	6.5	3.5	12.25	.168
1919	11,411	- 607	- 5.1	- 8.1	65.61	- .388
1920	13,429	2,018	17.7	14.7	216.09	.704
1921	7,945	- 5,484	- 40.8	- 43.8	1918.44	-2.099
1922	9,755	1,810	22.8	19.8	392.04	.949
1923	10,140	385	3.9	.9	.81	.043
1924	13,630	3,490	34.4	31.4	985.96	1.505
1925	16,105	2,475	18.2	15.2	231.04	.728
1926	17,978	1,873	11.6	8.6	73.96	.412
1927	12,956	- 5,022	- 27.9	- 30.9	954.81	-1.481
Total	---	---	---	---	7406.37	---

Table 2--Continued

7	8	9	10	11	12	13
Price						
Deflated seasonal average price in cents per pound 2/	First difference, or deviation from preceding year, in cents per pound	Percentage change (mean = 7.0)	Deviation of percentage change from mean	Square of deviation	Multiple of standard deviation (S.D. = 37.03)	Product of multiples of standard deviation
-	-	-	-	-	y	xy
13.68	---	---	---	---	---	---
10.11	- 3.57	-26.1	-33.1	1095.61	- .894	-1.379
11.49	1.38	13.6	6.6	43.56	.178	- .134
12.39	.90	7.8	.8	.64	.022	---
7.29	- 5.10	-41.2	-48.2	2323.24	-1.302	- .673
11.45	4.16	57.1	50.1	2510.01	1.353	-2.185
14.69	3.24	28.3	21.3	453.69	.575	- .014
15.50	.81	5.5	- 1.5	2.25	- .041	.009
14.32	- 1.18	- 7.6	-14.6	213.16	- .394	- .066
16.62	2.30	16.1	9.1	82.81	.246	- .095
7.55	- 9.07	-54.6	-61.6	3794.56	-1.664	-1.171
13.61	6.06	80.3	73.3	5372.89	1.979	-4.154
17.33	3.72	27.3	20.3	412.09	.548	.520
20.20	2.87	16.6	9.6	92.16	.259	.011
16.02	- 4.18	-20.7	-27.7	767.29	- .748	-1.126
12.56	- 3.46	-21.6	-28.6	817.96	- .772	- .562
8.60	- 3.96	-31.5	-38.5	1482.25	-1.040	- .428
14.53	5.93	69.0	62.0	3844.00	1.674	-2.479
Total	---	---	---	23308.17	---	-13.926

Table 2--Continued

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95. Production is in terms of 500-pound gross-weight bales.
2/ Adapted from column 3 of table 1.

Standard deviation of production = 20.87

Standard deviation of price = 37.03

Summation of products of corresponding multiples of standard deviation = -13.926

Coefficient of correlation (r) = $\frac{\Sigma xy}{n} = \frac{-13.926}{17} = -.819$

Coefficient of regression of price on production (b)

$= r\frac{\sigma y}{\sigma x} = -.817 \ \frac{37.03}{20.87} = -1.453.$

Table 3.- Deflated seasonal average price in cents per pound received by farmers for cotton lint and acreage of cotton harvested from subsequent plantings, 1910-27

Year	1	2	3	4	5	6
	Price					
	Deflated seasonal average price in cents per pound 1/	First difference, or deviation from preceding year, in cents per pound	Percentage change (mean = 7.0)	Deviation of percentage change from mean	Square of deviation	Multiple of standard deviation (S. D. = 37.03)
	-	-	-	-	-	x
1910	13.68	---	---	---	---	---
1911	10.11	-3.57	-26.1	-33.1	1095.61	- .894
1912	11.49	1.38	13.6	6.6	43.56	.178
1913	12.39	.90	7.8	.8	.64	.022
1914	7.29	-5.10	-41.2	-48.2	2323.24	-1.302
1915	11.45	4.16	57.1	50.1	2510.01	1.353
1916	14.69	3.24	28.3	21.3	453.69	.575
1917	15.50	.81	5.5	- 1.5	2.25	- .041
1918	14.32	-1.18	- 7.6	-14.6	213.6	- .394
1919	16.62	2.30	16.1	9.1	82.81	.246
1920	7.55	-9.07	-54.6	-61.6	3794.56	-1.664
1921	13.61	6.06	80.3	73.3	5372.89	1.979
1922	17.33	3.72	27.3	20.3	412.09	.548
1923	20.20	2.87	16.6	9.6	92.16	.259
1924	16.02	-4.18	-20.7	-27.7	767.29	- .748
1925	12.56	-3.46	-21.6	-28.6	817.96	- .772
1926	8.60	-3.96	-31.5	-38.5	1482.25	-1.040
1927	14.53	5.93	69.0	62.0	3844.00	1.674
Total	---	---	---	---	23308.17	---

Table 3--Continued

7	8	9	10	11	12	13
Acreage harvested						
Harvested acreage in thousands of acres reported for year following specified year 2/	First difference, or deviation from preceding item, in thousands of acres	Percentage change (mean = 1.7)	Deviation of percentage change from mean	Square of deviation	Multiple of standard deviation (S.D. = 10.0)	Product of multiples of standard deviation
-	-	-	-	-	y	xy
34,916	---	---	---	---	---	---
32,557	-2,359	- 6.8	- 8.5	72.25	- .850	.760
35,206	2,649	8.1	6.4	40.96	.640	.114
35,615	409	1.2	- .5	.25	- .050	- .001
29,951	-5,664	-15.9	-17.6	309.76	-1.760	2.292
33,071	3,120	10.4	8.7	75.69	.870	1.177
32,245	- 826	- 2.5	- 4.2	17.64	- .420	- .242
35,038	2,793	8.7	7.0	49.00	.700	- .029
32,906	-2,132	- 6.1	- 7.8	60.84	- .780	.307
34,408	1,502	4.6	2.9	8.41	.290	.071
28,678	-5,730	-16.7	-18.4	338.56	-1.840	3.062
31,361	2,683	9.4	7.7	59.29	.770	1.524
35,550	4,189	13.4	11.7	136.89	1.170	.641
39,501	3,951	11.1	9.4	88.36	.940	.243
44,386	4,885	12.4	10.7	114.49	1.070	- .800
44,608	222	.5	- 1.2	1.44	- .120	.093
38,342	-6,266	-14.0	-15.7	246.49	-1.570	1.633
42,434	4,092	10.7	9.0	81.00	.900	1.507
Total	---	---	---	1701.32	---	12.352

Table 3--Continued

1/ Adapted from column 3 of table 1.

2/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95. Statistics on acreages harvested are for the years 1911-28; 1911 acreage is paired with 1910 price, 1912 acreage is paired with 1911 price, etc. Statistics on acreage harvested are shown for specified years in column 1 of table 6.

Standard deviation of price = 37.03

Standard deviation of acreage = 10.00

Summation of products of corresponding multiples of standard deviation = 12.433

Coefficient of correlation (r) = $\frac{\Sigma xy}{n} = \frac{12.352}{17} = .727$

Coefficient of regression of acreage on price (b)

$= r\frac{\sigma y}{\sigma x} = .727\,\frac{10.0}{37.03} = .196.$

Estimates of percentage change in price can be made by making the proper substitutions in the formula y = Ay - bAx + bx, in which

y = the estimated percentage change in price

Ay = average of percentage changes in price

b = the coefficient of regression of price on production

x = the percentage change in production from one year to another

Ax = the average of percentage changes in production

Estimates of percentage change in price made by use of this formula are shown in column 2 of table 4. That part of the formula expressed as Ay - bAx becomes a constant, to which is added the product of b and x to obtain the estimate of percentage change in price for a given year.

Column 9 of table 2 shows, in the heading, that the value of Ay, the average of the percentage changes in price, is 7.0. The coefficient of regression is -1.453, and the average of percentage changes in production is 3.0. The average of the percentage changes in production is shown in the heading of column 3 of table 2. The constant is determined by subtracting from 7.0 (which is the value of Ay) the product of b (which is the coefficient of regression) and Ax (the average of the percentage changes in production). Since b, -1.453 is a minus quantity, the product of it and Ax is added to Ay to obtain the constant. We have, therefore, -1.453 times 3.0, which equals -4.4. When -4.4 is subtracted from 7.0 the constant becomes 11.4. The desired estimates, then, can be made by adding to 11.4 the product of b and x.

The estimates of percentage changes in price are shown in column 2 of table 4. Calculation of the estimate for 1911 will be explained to clarify

the procedure. Column 3 of table 2 shows that the production of 1911 was 35.2 percent greater than the production reported for 1910. The product of 35.2, which is x, and b, which is -1.453, is -51.1. The sum of -51.1 and 11.4, the constant, is -39.7, which is the estimate of the percentage change in price for 1911.

The estimating formula is not applied in this instance primarily for the purpose of showing the extent to which estimates of price changes approximate actual changes, but rather to illustrate the calculations that are necessary when it can be applied advantageously. The facts are that in many instances the estimates in table 4 differ so greatly from the actual changes that the formula as described is not particularly applicable, indicating that effects of factors other than production on acreage harvested might need to be taken into consideration to obtain the desired approximations of estimated price changes. However, the calculations illustrate the substitutions that are to be made in the formula when it is used.

In table 2, production and price of the same year have been correlated, the assumption being that some indication of the probable size of the crop will be available before any considerable part of it has been ginned. Some improvement in the estimates might be made by correlating production with the average of prices for the months during which farmers market most of their crop. A suggestion to the instructor is that students be assigned to make the estimates by correlating production with prices of September, October, November, and December and then proceeding by the method described. A further suggestion is that undeflated prices be used also in determining the relationship between production and price in order to furnish a comparison between estimates obtainable by calculations into which deflated prices enter.

Table 4.- Actual and estimated percentage change in deflated seasonal average price of cotton, 1911-27

Year	1 Actual percentage change 1/	2 Estimated percentage change 2/
	Percent	Percent
1911	- 26.1	- 39.7
1912	13.6	29.9
1913	7.8	6.6
1914	- 41.2	- 8.7
1915	57.1	56.0
1916	28.3	7.8
1917	5.5	13.4
1918	- 7.6	2.0
1919	16.1	18.8
1920	- 54.6	- 14.3
1921	80.3	70.7
1922	27.3	- 21.7
1923	16.6	5.7
1924	- 20.7	- 38.6
1925	- 21.6	- 15.0
1926	- 31.5	- 5.5
1927	69.0	51.9
Average	7.0	7.0

1/ Adapted from column 9 of table 2.
2/ Calculated by the formula y = Ay - bAx + bx, in which y = the estimated percentage change in price; Ay, the average of percentage changes in price; b, the coefficient of regression of price on production; x, the percentage change in production; and Ax, the average of percentage changes in production.

Column 1 of table 5 shows the actual percentage changes in acreage of cotton harvested, whereas column 2 shows the estimated percentage changes. The estimates were made by making the proper substitutions in the formula $y = Ay - bAx + bx$, in which

y = the estimated percentage change in acreage harvested
Ay = the average of percentage changes in acreage harvested

b = the coefficient of regression of acreage on price

x = the percentage change in price from one year to another

Ax = the average of percentage changes in price.

The constant in the formula is $Ay - bAx$, to which is added the product of b and x. The value of Ay is 1.7, shown in the heading of column 9 of table 3 ; the value of b is .196; and the value of Ax is 7.0. To obtain the value of the constant we subtract from 1.7 the product of .196 and 7.0, which leaves .3. Estimates in column 2 of table 5 can now be made by adding to .3 the product of b and x. The calculations will be illustrated by making the estimate for 1912.

Table 3 shows that the price for 1911 was 26.1 percent less than the price for 1910. The product of -26.1 and .196 is -5.1. The sum of -5.1 and .3, the constant, is -4.8, the estimated percentage change in acreage for 1912. Estimated percentage changes in acreages harvested from plantings of other years are determined in the same manner, by making the proper substitutions in the formula.

In this instance also the primary purpose of making the estimates is not to show how close the actual changes in acreage harvested can be approximated by use of the formula, but rather to illustrate the procedure.

Table 5.- Actual and estimated percentage change in acreage of cotton harvested in the United States, 1912-28

Year	1 Actual percentage change 1/	2 Estimated percentage change
	Percent	Percent
1912	-6.8	-4.8
1913	8.1	3.0
1914	1.2	1.8
1915	-15.9	-7.8
1916	10.4	11.5
1917	-2.5	5.8
1918	8.7	1.4
1919	-6.1	-1.2
1920	4.6	3.5
1921	-16.7	-10.4
1922	9.4	16.0
1923	13.4	5.7
1924	11.1	3.6
1925	12.4	-3.8
1926	.5	-3.9
1927	-14.0	-5.9
1928	10.7	13.8
Average	1.7	1.7

1/ Adapted from column 9 of table 3. See footnote 2 of that table. Statistics on acreage harvested are shown in column 1 of table 6.

Ratio Method of Estimating 1/

In some instances it might be possible to make fairly dependable estimates by the ratio method, thus obviating the necessity of calculating the coefficient of correlation and the coefficient of regression. By the ratio method it is only necessary to express the independent variables in terms of ratios, obtain their product, and then multiply the product by the magnitude of the dependent variable preceding the magnitude to be estimated.

To illustrate the procedure, estimates of cotton production in the United States will be made for the period 1900-36. In order to clarify the computations all of the statistical information used is included.

Table 6 shows statistics on the acreage and production of cotton for the period 1891-1936. The annual average yield of lint cotton per acre is shown by the statistics in column 1 of table 7, and 9-year average yields are shown in column 2 of that table. These 9-year average yields were calculated from table 6. Statistics on the actual production of lint cotton are shown in table 8. Yield ratios are shown in column 1 of table 9. These ratios were calculated by dividing the 9-year average yields per acre in column 2 of table 7 by the annual average yields per acre in column 1 of that table. Acreage ratios are shown in column 2 of table 9. These ratios were calculated by dividing the acreages of specified years by acreages of

---

1/ Acknowledgment is made of suggestions received from Bradford B. Smith of the Cleveland Trust Company, Cleveland, Ohio, formerly associated with the United States Department of Agriculture.

[illegible]

preceding years, as explained in footnote 2 of table 9. Tables 10 and 11 show statistics on actual production, the estimates of production, and the extent to which estimated production differs from actual. Footnote 2 of table 10 explains the procedure by which the estimates in that table were made, and footnote 3 of table 11 explains how the estimates in that table were made.

Table 6.- Acreage and production of cotton in the United States, 1891-1936

Year	1 Acreage harvested 1/	2 Production of lint cotton 1/
	1,000 acres	1,000 bales
1891	21,503	9,035
1892	18,869	6,700
1893	20,256	7,493
1894	21,886	9,901
1895	19,839	7,162
1896	23,230	8,533
1897	25,131	10,899
1898	24,715	11,278
1899	24,163	9,346
1900	24,886	10,124
1901	27,050	9,508
1902	27,561	10,630
1903	27,762	9,851
1904	30,077	13,438
1905	27,753	10,576
1906	31,404	13,274
1907	30,729	11,106
1908	31,091	13,241
1909	30,555	10,005
1910	31,508	11,609
1911	34,916	15,694

Table 6.- Continued

Year	1 Acreage harvested 1/	2 Production of lint cotton 1/
	1,000 acres	1,000 bales
1912	32,557	13,703
1913	35,206	14,153
1914	35,615	16,112
1915	29,951	11,172
1916	33,071	11,448
1917	32,245	11,284
1918	35,038	12,018
1919	32,906	11,411
1920	34,408	13,429
1921	28,678	7,945
1922	31,361	9,755
1923	35,550	10,140
1924	39,501	13,630
1925	44,386	16,105
1926	44,608	17,978
1927	38,342	12,956
1928	42,434	14,477
1929	43,232	14,825
1930	42,444	13,932
1931	38,704	17,097
1932	35,891	13,003
1933	29,383	13,047

Table 6.- Continued

Year	1 Acreage harvested 1/	2 Production of lint cotton 1/
	1,000 acres	1,000 bales
1934	26,866	9,636
1935	27,640	10,638
1936	30,028	12,399

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95. Production is in terms of running bales for the years 1891-98, and in terms of 500-pound gross-weight bales for the years 1899-1936.

Table 7.- Annual and 9-year average yield of cotton lint per acre in the United States, 1899-1936

Year	1 Annual average yield of lint cotton per acre 1/	2 Average yield of lint cotton per acre for 9-year period ending with specified year 2/
	Pounds	Pounds
1899	185.0	192
1900	194.7	192
1901	168.2	191
1902	184.7	191
1903	169.9	186
1904	213.7	191
1905	182.3	191
1906	202.3	191
1907	172.9	186
1908	203.8	188
1909	156.5	184
1910	176.2	185
1911	215.0	189
1912	201.4	192
1913	192.3	190
1914	216.4	194
1915	178.5	191
1916	165.6	190
1917	167.4	186
1918	164.1	187

Table 7.- Continued

Year	1	2
	Pounds	Pounds
1919	165.9	185
1920	186.7	182
1921	132.5	175
1922	148.8	170
1923	136.4	161
1924	165.0	160
1925	173.5	161
1926	192.9	165
1927	161.7	164
1928	163.3	164
1929	164.2	162
1930	157.1	164
1931	211.5	170
1932	173.5	173
1933	212.7	177
1934	171.6	177
1935	184.2	176
1936	197.6	180

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95.

2/ Calculated from table 6. Average yields per acre in parts of bales converted to average yields in pounds on the basis of 478 pounds, net, of lint per bale.

Table 8.- Production of cotton in the United States, 1899-1936

Year	Production of lint cotton 1/
	1,000 bales
1899	9,346
1900	10,124
1901	9,508
1902	10,630
1903	9,851
1904	13,438
1905	10,576
1906	13,274
1907	11,106
1908	13,241
1909	10,005
1910	11,609
1911	15,694
1912	13,703
1913	14,153
1914	16,112
1915	11,172
1916	11,448
1917	11,284
1918	12,018
1919	11,411
1920	13,429

Table 8.- Continued

Year	Production of lint cotton 1/
	1,000 bales
1921	7,945
1922	9,755
1923	10,140
1924	13,630
1925	16,105
1926	17,978
1927	12,956
1928	14,477
1929	14,825
1930	13,932
1931	17,097
1932	13,003
1933	13,047
1934	9,636
1935	10,638
1936	12,399

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95. Production is in terms of 500-pound gross-weight bales.

Table 9.- Yield and acreage ratios, 1899-1936

Year	Yield ratio 1/	Acreage ratio 2/
1899	1.038	.978
1900	.986	1.030
1901	1.136	1.087
1902	1.034	1.019
1903	1.095	1.007
1904	.894	1.083
1905	1.048	.923
1906	.944	1.132
1907	1.076	.979
1908	.922	1.012
1909	1.176	.983
1910	1.050	1.031
1911	.879	1.108
1912	.953	.932
1913	.988	1.081
1914	.896	1.012
1915	1.070	.841
1916	1.147	1.104
1917	1.111	.975
1918	1.140	1.087
1919	1.115	.939
1920	.975	1.046
1921	1.321	.833

Table 9.- Continued

Year	Yield ratio 1/	Acreage ratio 2/
1922	1.142	1.094
1923	1.180	1.134
1924	.970	1.111
1925	.928	1.124
1926	.855	1.005
1927	1.014	.860
1928	1.004	1.107
1929	.987	1.019
1930	1.044	.982
1931	.804	.912
1932	.997	.927
1933	.832	.819
1934	1.031	.914
1935	.955	1.029
1936	.911	1.086

1/ Nine-year average yield in column 2 of table 7 divided by annual average yield in column 1 of table 7.

2/ Acreage of specified year in table 6 divided by acreage of preceding year.

Table 10.- Actual and estimated production of cotton in the United States, 1899-1936

Year	1 Actual production of lint 1/	2 Estimated production of lint 2/	3 Extent to which estimated production deviates from actual production
	1,000 bales	1,000 bales	1,000 bales
1899	9,346	-	-
1900	10,124	9,488	- 636
1901	9,508	10,282	774
1902	10,630	11,741	1,111
1903	9,851	11,200	1,349
1904	13,438	10,862	-2,576
1905	10,576	13,011	2,435
1906	13,274	10,230	-3,044
1907	11,106	14,185	3,079
1908	13,241	11,699	-1,542
1909	10,005	12,355	2,350
1910	11,609	11,566	- 43
1911	15,694	12,567	-3,127
1912	13,703	15,285	1,582
1913	14,153	12,171	-1,982
1914	16,112	15,116	- 996
1915	11,172	14,610	3,438
1916	11,448	10,053	-1,395
1917	11,284	14,496	3,212
1918	12,018	12,223	205

Table 10.- Continued

Year	1	2	3
	1,000 bales	1,000 bales	1,000 bales
1919	11,411	14,892	3,481
1920	13,429	11,947	-1,482
1921	7,945	13,696	5,751
1922	9,755	8,743	-1,012
1923	10,140	12,187	2,047
1924	13,630	13,569	- 61
1925	16,105	14,689	-1,416
1926	17,978	16,799	-1,179
1927	12,956	15,448	2,492
1928	14,477	11,298	-3,179
1929	14,825	16,090	1,265
1930	13,932	14,910	978
1931	17,097	14,283	-2,814
1932	13,003	12,536	- 467
1933	13,047	12,018	-1,029
1934	9,636	8,890	- 746
1935	10,638	9,080	-1,558
1936	12,399	10,454	-1,945
Average	-	-	22

1/ U. S. Department of Agriculture. Agricultural Statistics, 1938, table 114, pages 94 and 95. Production is in terms of 500-pound gross-weight bales.

2/ Production of preceding year times the product of yield and acre ratios in table 9. For example, production of 1899 is multiplied by the product of the yield and acreage ratios shown opposite 1899 in table 9 to obtain the estimate of production for 1900.

Table 11.- Actual and estimated production of cotton in the United States, 1899-1936

Year	1 Average production for 3-year period ending with specified year 1/	2 Production of lint 2/	3 Estimated production of lint 3/	4 Extent to which estimated production deviates from actual production
	1,000 bales	1,000 bales	1,000 bales	1,000 bales
1899	10,508	9,346	-	-
1900	10,249	10,124	10,667	543
1901	9,659	9,508	10,409	901
1902	10,087	10,630	11,927	1,297
1903	9,996	9,851	10,628	777
1904	11,306	13,438	11,022	-2,416
1905	11,288	10,576	10,946	370
1906	12,429	13,274	10,919	-2,355
1907	11,652	11,106	13,282	2,176
1908	12,540	13,241	12,274	- 967
1909	11,451	10,005	11,701	1,696
1910	11,618	11,609	13,237	1,628
1911	12,436	15,694	12,577	-3,117
1912	13,669	13,703	12,112	-1,591
1913	14,517	14,153	12,141	-2,012
1914	14,656	16,112	15,505	- 607
1915	13,812	11,172	13,289	2,117
1916	12,911	11,448	12,429	981
1917	11,301	11,284	16,349	5,065

Table 11.- Continued

Year	1	2	3	4
	1,000 bales	1,000 bales	1,000 bales	1,000 bales
1918	11,583	12,018	12,242	224
1919	11,571	11,411	14,353	2,942
1920	12,286	13,429	12,115	-1,314
1921	10,928	7,945	12,530	4,585
1922	10,376	9,755	12,025	2,270
1923	9,280	10,140	12,963	2,823
1924	11,175	13,630	12,418	-1,212
1925	13,292	16,105	12,043	-4,062
1926	15,904	17,978	13,865	-4,113
1927	15,680	12,956	13,666	710
1928	15,137	14,477	13,674	- 803
1929	14,086	14,825	16,824	1,999
1930	14,411	13,932	14,167	235
1931	15,285	17,097	14,774	-2,323
1932	14,677	13,003	11,208	-1,795
1933	14,382	13,047	13,565	518
1934	11,895	9,636	9,800	164
1935	11,107	10,638	11,209	571
1936	10,891	12,399	10,915	-1,484
Average	-	-	-	65

1/ Calculated from table 6.

2/ Adapted from table 6.

3/ Three-year average production in column 1 times the product of yield and acreage ratios in table 9. For example, 3-year average production shown opposite 1899 is multiplied by the product of yield and acreage ratios shown opposite 1899 in table 9 to obtain the estimate of production for 1900.

Government Forecasts of Cotton Production

Among the forecasts made by the United States Department of Agriculture are those on the cotton crop. Information on the making of these estimates is contained in United States Department of Agriculture Miscellaneous Publication Number 171, published in November, 1933. Table 12 is included to provide a comparison between the Government's forecasts of the size of the cotton crop and the actual size of the crop as officially reported by the Bureau of the Census.

Table 12.- Government forecasts and estimates of the cotton crop and size of the crop as officially reported after the end of the ginning season, 1915-37

Year	Forecast 1/ As of July 25	As of Aug. 25	As of Sept. 25	As of Oct. 25	Estimate as of Dec. 1 1/	Actual ginnings 2/
	1,000 bales	1,000 bales	1,000 bales	1,000 bales	1,000 bales	1,000 bales
1915	11,876	11,697	10,950	-	11,161	11,192
1916	12,916	11,800	11,637	-	11,511	11,450
1917	11,949	12,499	12,047	-	10,949	11,302
1918	13,619	11,137	11,818	-	11,700	12,041
1919	11,016	11,230	10,696	-	11,030	11,421
1920	12,519	12,783	12,123	-	12,987	13,440
1921	8,203	7,037	6,537	-	8,340	7,954
1922	11,447	10,575	10,135	-	9,964	9,762
1923	11,517	10,788	11,015	10,248	10,081	10,140
	As of Aug. 1	As of Sept. 1	As of Oct. 1	As of Nov. 1	As of Dec. 1	Actual ginnings
1924	12,351	12,787	12,499	12,816	13,153	13,628
1925	13,566	13,740	14,759	15,386	15,603	16,104
1926	15,621	15,166	16,627	17,918	18,618	17,977
1927	13,492	12,692	12,678	12,842	12,789	12,956
1928	14,291	14,439	13,993	14,133	14,373	14,478
1929	15,543	14,825	14,915	15,009	14,919	14,825
1930	14,362	14,340	14,486	14,438	14,243	13,932
1931	15,584	15,685	16,284	16,903	16,918	17,096
1932	11,306	11,310	11,425	11,947	12,727	13,002
1933	12,314	12,414	12,885	13,100	13,177	13,047
1934	9,195	9,252	9,443	9,634	9,731	9,637

Table 12.- Continued

Year	Forecast 1/ As of Aug. 1	As of Sept. 1	As of Oct. 1	As of Nov. 1	Estimate as of Dec. 1 1/	Actual ginnings 2/
	1,000 bales	1,000 bales	1,000 bales	1,000 bales	1,000 bales	1,000 bales
1935	11,798	11,489	11,464	11,141	10,734	10,638
1936	12,481	11,121	11,609	12,400	12,407	12,399
1937	15,593	16,098	17,573	18,243	18,746	18,945

1/ Statistics assembled from a mimeographed report of the United States Department of Agriculture dated June 15, 1938, and entitled "Comparison of Cotton Forecasts, Estimates, and Final Ginnings." Forecasts and estimates are in terms of 500-pound gross-weight bales.

2/ U. S. Department of Commerce. Bureau of the Census. Bulletin 175, table 3, page 4. Ginnings are in terms of 500-pound gross-weight bales.

Standard Error of Estimate

Frequently when estimates are made it is desirable to measure the degree of approach to absolute accuracy in order to evaluate the estimating procedure. In using the forecasting formula herein described the approach to accuracy in the estimates is generally expected to vary in accordance with the magnitude of the coefficient of correlation, but this is not always the case because in determining the coefficient of regression the standard deviations have an effect. Furthermore, abnormally large or small items in either of the series correlated may so greatly affect the coefficient of correlation that the calculated measure of correlation is not a true indication of normal cause and effect relationships. This is a circumstance that often must be reckoned with in correlation procedure, and formulas in which the coefficient of regression is used might sometimes be unsatisfactory for estimating purposes in spite of any magnitude of the coefficient of correlation that might be obtained from the analysis of relationships.

The standard error of estimate is merely a numerical quantity that serves as an indication of the dependability of the estimated values. It is calculated by extracting the square root of the mean-square-deviation of estimated values from actual values. First, the deviations of estimated values from actual values are determined. Then these deviations are squared, the products are summated, the sum is divided by the number of deviations, and the square root of the quotient is extracted. The standard error so calculated is generally interpreted as an indication that if a distance equal to the standard error is measured from both sides of the arithmetic mean of estimates about 68.26 percent of all the estimated values will be included. Of course, this is not necessarily true any more than a distance

equal to the standard deviation measured from both sides of the arithmetic mean of distributions that are not normal will include about 68.26 percent of all the items. It is commonly stated that when the standard error of estimate is greater than the standard deviation of actual values the estimating formula cannot be relied upon.

When actual values are correlated with estimated values it should be possible to obtain a coefficient of correlation of about the same magnitude as the coefficient of correlation used in constructing the estimating formula. That is, all or most of the original coefficient of correlation should be recovered when actual values are correlated with estimated values.

Coefficient of Alienation

The coefficient of alienation is a measure of the absence of correlation just as the coefficient of correlation measures the presence of correlation. It is determined by comparing the standard error of estimate (or the standard deviation of the differences between estimated and actual values) with the standard deviation of actual values.

The coefficient of alienation can be calculated by dividing the standard error of estimate by the standard deviation of the actual values. If we divide the square of the standard error of estimate by the square of the standard deviation of actual values we obtain the coefficient of non-determination, or the square of the coefficient of alienation. For example, let us assume that the square of the standard error of estimate is 9 and that the square of the standard deviation of the actual values is 16. The quotient obtained by dividing 9 by 16 is .5625, the coefficient of non-determination, or the square of the coefficient of alinenation. The square root of .5625 is .75,

which is the coefficient of alienation. The coefficient of alienation is obtainable also by dividing the standard error of estimate (square root of 9) by the standard deviation of actual values (square root of 16).

The coefficient of determination is the square of the coefficient of correlation.