Predicting Sugarbeet Storage Losses Using Regression Analysis

M. G. BARNES, W. R. AKESON and N. PENCE[1]

Received for publication April 10, 1974

Introduction

Accurate prediction of weight and sugar losses of stored sugarbeets at the beginning of the storage period is useful because it enables the total losses to be estimated by the end of harvest, and it helps to explain the causes of such losses.

Regression analysis has been used previously in agricultural prediction; for example, to predict the yield of corn (4)[2] and of crested wheatgrass (2), and to analyze relationships between yield and weather in sugarbeets (1). However, no reference was found for the use of this method to predict actual storage losses.

Methods and Materials

Calculations were made using the Burroughs ASSIST statistical-program package (3). The weather data is from *CLIMATOLIGICAL DATA* (5), and all other data is from records of The Great Western Sugar Company. Data from 1969 were eliminated because they are not comparable to other years' data due to frozen beets.

Multiple linear regression analysis was used to find those independent variables which best explain the historical losses. These factors were considered for use as independent variables:

1. Campaign length.
2. Average number of storage days.
3. Deviation from normal temperatures for weekly periods in October.
4. Weekly maximum and minimum temperatures in October.
5. Percent of beets delivered during weekly periods in October.
6. Percent of beets delivered by October 8, 16, or 27.
7. Precipitation during the period September 15 to October 31.
8. Percent of beets piled after first occurrence of 24°F. or below, and also 20°F. or below.
9. Deviation from normal of average monthly temperatures in November and December.

---

[1]Data Analyst, Sr. Plant Physiologist, and Operation Research Analyst, respectively, The Great Western Sugar Company, Agricultural Research Center, Longmont, Colorado 80501.

[2]Numbers in parentheses refer to literature cited.

Each geographical region was analyzed separately with the same technique. Detailed results are given for one region, ,North Central Colorado, which includes the Eaton, Greeley, Loveland, Longmont, and Brighton factory districts. Weather data for this region are from the Greeley and Longmont 2ESE stations.

Factors were eliminated if their correlation with the shrink was of magnitude less than 0.3, or because two factors showed correlation of magnitude 0.3 or more with each other and could not be used in the same equation.

Factors of the first type were eliminated initially. If no more than about eight factors remained, equations were generated by using each variable with as many of the other variables as possible, in all combinations which did not include inter-correlated variables in the same equation. If more than eight factors remained, those which showed the highest correlation with the shrink were used to generate equations first and other variables added or substituted if necessary. Sometimes, statistical tests showed undesirable characteristics in these equations. In these cases, either one or more variables were dropped from the equations, or the whole set of variables was discarded.

After the elimination of variables had been completed, the equations were compared and the one with highest multiple R-square became the predictor equation. This method was used to find a weight shrink and a sugar loss equation for each region.

Results and Discussion

In North Central Colorado, the weight shrink equation variables are:

(1) Campaign length.
(2) Percent of beets delivered by October 16.
(3) Maximum temperature, October—Week 4.

The sugar loss equation uses these variables:

(1) Campaign length.
(2) Percent of beets delivered by October 16.
(3) Maximum temperature, October—Week 2.
(4) Maximum temperature, October—Week 4.

The most variables in any equation is four, the least is two. More factors actually affect storage losses, but these effects are overshadowed by random variability under field conditions, so only the strongest factors are useful in prediction. All of the equations for sugar loss use only three basic factors: campaign length, October temperatures (November temperatures for Ohio), and rate of delivery.

The values estimated from the equations for North Central Colorado are compared (Table 1) with actual values for the period 1960-1971 As in all regions, the sugar loss equation gives better results

Table 1.—Comparison of estimated with true values of shrink north central Colorado.

Year	Sugar Loss		Weight Loss	
	Actual	Estimate	Actual	Estimate
1971	129.13*	128.52*	210.53†	178.22†
1970	97.51	100.95	91.19	76.60
1968	73.59	75.61	78.69	71.91
1967	88.59	93.86	48.46	62.53
1966	87.57	81.49	28.66	23.45
1965	59.80	59.80	2.61	-0.52
1964	108.25	105.21	162.58	144.87
1963	116.36	113.93	167.80	177.18
1962	93.86	100.34	118.29	150.08
1961	114.74	107.24	85.46	112.04
1960	130.55	133.18	105.78	103.18

*Percent of 11-year north central Colorado sugar loss mean.
†Percent of 11-year north central Colorado weight loss mean.

than the weight shrink equation, due mainly to variability in the method of measuring weight shrink.

A comparison of the multiple R-square and the standard error of estimate for the equations (Table 2) indicates their accuracy; the weight shrink equations account for between 74 and 91 percent of the variation in weight shrink, and the sugar loss equations account for 89 to 96 percent of the variation in sugar loss. Company-wide comparisons of predicted shrink values with the actual values are in Table 3.

Table 2.—Correlation coefficients and standard error of estimates of multiple regression equations.

Region	Multiple R-Square		Standard Error of Estimate	
	Weight loss	Sugar Loss	Weight Loss	Sugar Loss
N. C. Colorado	0.91	0.96	18.83*	5.44†
Northeast Colorado	0.74	0.89	22.68	7.19
Nebraska	0.91	0.94	20.11	8.17
Wyoming	0.83	0.95	37.66	11.08
Montana	0.79	0.91	32.10	19.25
Northern Ohio	0.81	0.89	35.95	21.97

*Percent of 9-year company-wide weight loss mean.
†Percent of 9-year company-wide sugar loss mean.

The final test of the usefulness of the equations is how storage losses predicted in advance compare with the actual losses (Table 4). The first projection was made by November 5, 1972 (December 5 for Ohio) using an estimate of campaign length, and the second was made later using the actual campaign length. The accuracy of these predictions shows the great value of this technique.

Summary

A study was carried out to determine factors to use in multiple linear regression equations for predicting weight and sugar shrink in

Table 3.—Comparison of estimated with true values of shrink, company-wide.

Year	Sugar* Loss		Weight† Loss	
	Actual	Estimate	Actual	Estimate
1971	142.52	149.13	160.06	161.35
1970	104.99	107.13	104.85	102.29
1968	92.16	93.13	90.73	91.59
1967	97.80	94.88	86.45	88.16
1966	93.91	90.41	65.91	71.90
1965	74.85	79.72	33.81	40.66
1964	96.82	95.46	119.83	112.56
1963	122.10	116.07	154.93	143.80
1962	74.85	80.30	83.45	87.73

*Percent of 9-year company-wide sugar loss mean.
†Percent of 9-year company-wide weight loss mean.

Table 4.—Comparison of 1972 projections with actual shrink values.

	N. Central Colo.	Company-wide
Sugar Loss*		
1st Projection ‡	125.99	136.63
2nd Projection §	127.54	140.42
Actual	145.82	152.46
Weight Loss: †		
1st Projection ‡	188.74	163.61
2nd Projection §	196.44	166.46
Actual	121.12	148.21

* Percent of 9-year company-wide sugar loss mean.
†Percent of 9-year company-wide weight loss mean.
‡1st projection made on or before Nov. 5, 1972 using estimated campaign length.
§2nd projection made at end of campaign, using actual campaign length.

stored sugarbeets. Geographic regions were analyzed separately. Resulting weight shrink equations have R-square values ranging from .74 to .91 and sugar loss equations have R-square values between .89 and .96. Predictions using the equations made at the end of harvest in 1972 were close to actual shrink values.

Literature Cited

(1) BRUMMER, VEIKKO. 1961. On the relations between sugarbeet yields and certain climatic factors in Finland. A statistical study and some cultivation technique applications. (In Finnish) Acta. Agralia Fennica 98:7-180.

(2) CURRIE, PAT O. and GERALDINE PETERSON. 1966. Using growing-season precipitation to predict crested wheatgrass yields. J. Range Management 19:284-288.

(3) DATA SERVICES DEPARTMENT, FEDERAL BANK OF BOSTON. 1970. ASSIST-B2500/B3500 Statistical System, and User Manual. Burroughs Corporation.

(4) RUNGE, EDWARD C. A. 1970. Use weather to predict your corn yields. Crops and Soils 22(5):11-13.

(5) U.S. DEPARTMENT OF COMMERCE. 1950-1972. Climatological Data. National Climatic Center, Asheville, N.C. 28801.