Design and Analysis of Onion Storage Trials[1]

By Robert L. Plaisted, T. W. Horner, and E. P. Lana,
Iowa State College, Ames, Iowa[2]

Onion breeders have been interested in a number of problems which have arisen in the conduct of onion storage trials. One of these problems is that the data are not available until spring, —a fact which can complicate advance planning for the breeding plots. Other problems are estimates of the number of replications needed to detect differences of a given magnitude, effect of sample size, storage location, and length of storage, and finally whether the data should be analyzed as a percentage of loss or transformed to angles by the arc sine transformation.

In connection with the first problem, Woodman and Barnell (11) investigated the possibility of comparing varieties by measuring water loss over a short period of time. They felt that although there might be difficulty in spacing the varieties in any ranked order of keeping quality, they could at least separate the varieties into two groups, keepers and non-keepers, based on total water loss and relative rate of loss.

The need for the arc sine transformation is dependent upon how much better the transformed data satisfy the assumptions underlying the analysis of variance than the percentage data. Eisenhart (7) has defined these assumptions and Cochran (6) discussed the consequences when they are not satisfied. The use of the transformation has been discussed in the general case by Bliss (1, 2), Clark and Leonard (3), and Cochran (4, 5).

Materials and Methods

The storage used in 1953–54 and one of the storage locations used in 1954–55 was the modified common storage designed and built for the onion breeding project at Ames, Iowa. This storage has no mechanical refrigeration but is equipped with a blower to circulate the air inside the storage and to introduce cooler air from the outside when it is available and is needed to obtain and maintain a temperature of 32 degrees F. There are three cone heaters connected to a thermostat which prevent temperatures inside the storage dropping below the freezing temperature. The second storage location in the 1954–55 trial was in the basement of the horticulture building where the temperature was usually near 80 degrees Fahrenheit, and the relative humidity about 20 per cent.

The bulbs were grown on peat soil at Clear Lake and Fertile, Iowa. In 1953 these were Brigham Yellow Globe, 2190 x 2108, 2264 x Iowa 163, and 2271 x 2215. In 1954 the varieties were Asgrow Y-42, Condon's Autumn Spice, Asgrow Y-51, and Rochester Bronze. Samples

[1]Received for publication July 30, 1956. Journal Paper No. J-2995 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project No. 1129.

[2]Present addresses: Robert L. Plaisted, Cornell University, Ithaca, N.Y.; E. P. Lana, North Dakota Agricultural College, Fargo, N.D.

of 25, 50, and 100 bulbs were put in mesh bags, tied with tags identifying variety and sample size, weighed, and placed in the wire bottom storage trays. Three trays stacked one on top of another constituted a replicate. One of these trays had one 100-bulb sample for each variety, another had two 50-bulb samples for each variety, and the last had four 25-bulb samples for each variety. The two 50-bulb and the four 25-bulb samples were each arbitrarily given a suffix to identify them for recording purposes. Thus there are two and four independent estimates of the parameters involving 50-bulb and 25-bulb samples respectively. There were nine replications at each location in 1954–55 and six replications in 1953–54.

Storage data were taken in 1953–54 on March 18, April 10, and May 10; in 1954–55 on December 20, January 22, February 19, March 26, and April 23 in the onion storage and on December 21, January 21, February 18, and March 26 in the horticulture basement storage. The initial weight into storage for each sample had been recorded when the samples were placed in storage. On each date, data showing weight loss since the beginning of storage, and numbers of rots and sprouts were recorded.

Parallel analyses of variance of percentages and percentages transformed to degrees by the arc sine transformation (9) were run on all the data. In addition to comparisons of main effects, interactions, and relative efficiency of design when estimated by these two sets of data, it seemed worthwhile to investigate the matters of homogeneity of variance, distribution of errors, and additivity of treatment and environmental effects. The homogeneity of the variances within each of the four varieties within each sample size in each date as calculated on the basis of both percentages and angles was measured by chi-square with 3 degrees of freedom, as suggested by Bartlett and described in Snedecor (9). The test for non-additivity used is the one described by Tukey (10).

In order to study the extent to which error variation was binomial, the variances were plotted against their mean per cent loss. Since the variances of the percentage data appear to be somewhat dependent on the mean approximately similar to binomially distributed data, it is important to have some mathematical expression of the goodness of fit. This was done by assuming the following model under the alternative hypothesis:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Here Y_i = the observed variance

α = the intercept on the ordinate

X_i = $p_i(1-p_i)$

p_i = mean percentage loss

β = n' which corresponds to n in the expected binomial variance $np(1-p)$.

ϵ_i = the experimental error variation of the i^{th} observation.

The hypothesis that $\beta = n' = o$ was tested.

It should be noted that n' for these continuous data has no real meaning comparable to the n of discrete data, but operates as a scaler. Further, in fitting regressions, the X_i are considered as being fixed. This is not strictly true here. However, if we only wish to reach some conclusion about the nature of the data at hand, we may assume this requirement to be fulfilled.

The estimates of the number of replications required to detect a difference of a given magnitude with a prescribed level of probability are based on procedure and tables given by Neyman and Tokarska (8).

The effect of sample size and variety on rotting and sprouting was measured by use of chi-square as suggested by Snedecor (10) on page 210. In comparing the differences among and within sample sizes, the four 25-bulb and the two 50-bulb samples were totaled, so that in all cases, the numbers analysed are based on an original 100 bulbs. In evaluating the differences among sample sizes, the numbers lost were totaled over varieties and likewise when considering variety differences, the numbers lost were totaled over sample sizes.

Results

Effect of location.—It would be an advantage to the plant breeder if he could select a storage location which would accelerate storage losses and give the same information he had been getting from a more prolonged storage period under more favorable storage conditions. However, analyses of variance do not support the choice of such a procedure, primarily because of the highly significant variety-by-location interaction. This interaction is high, whether considering the two locations at a given date, or considering the two locations at two dates with the same mean percentage weight loss.

The effect of changing storages is not one of simple acceleration, but rather a complex one which changes the character of the storage curve for each variety. Further, the nature of this change is not the same for each variety. Thus it appears necessary to select a storage comparable to those in use by the growers for whom the hybrids are being developed. This is true whether one is making recommendations on the length of storage for any given variety, or comparing hybrids being developed to fit the requirements of these growers.

Length of storage period:—Another alternative in order to obtain storage information earlier would be merely to take the data sooner. Here again, this is complicated by a large variety-by-length-of-storage interaction. Thus the length of storage should be conditioned by the length of time the growers usually need to store their crop.

Size of sample:—Size of sample is important largely as it concerns efficiency of estimates, though the interactions of variety, location, and length of storage period with sample size are important in comparing results from different projects or in making interpretations of the data at hand adjustable to growers' conditions.

The significance of the main effect of sample size, using the word "main" in the sense used in factorial analyses, is certainly not conclusive from the data available. In 1954–55 in both locations, the smaller the sample size, the greater was the loss; but in 1953–54 in the onion storage, the larger the sample size the greater the loss with the exception of the last date. A few of these differences were significant. When the data were combined over locations at each date, the effects of sample size were small, but approaching significance. So, generally speaking, sample size apparently does have some direct effect on storage loss, but this effect is neither sizable nor consistent, though it would appear that the smaller samples tend to lose slightly more weight than do the larger.

The variety-by-sample-size interaction was also somewhat inconsistent. It is interesting to note that this interaction of sample size with variety was significant only in the 1954–55 onion storage, which was consistently non-significant for differences among the sample sizes. Even in these cases where the interaction was significant, the level of significance was low.

The sample-size-by-length-of-storage and sample-size-by-location interactions were small and non-significant. Likewise, the two second order interactions of length-of-storage-by-variety-by-sample-size and variety-by-sample-size-by-location were non-significant.

Relationship of the coefficient of variation to sample size and length of storage period:—The coefficients of variability for the three sample sizes at the several dates for both the percentage data and the transformed data are given in Table 1. There is a consistent decrease in the size of the coefficient of variability as the size of sample increases. Concurrent with this is a decrease in the coefficient of variation as the storage period is extended. The transformation decreased the coefficient of variation considerably, particularly for those dates and sample sizes where the coefficient of variation of the percentage data was large. It can be seen from Table 1 that there is a very close negative association between the size of the coefficient of variation and the over-all mean per cent loss. This is true whether one com-

Table 1.—Coefficients of variability of per cent and transformed per cent loss (expressed as a percentage).

	Per cent[a]				Transformed per cent[b]			
	Sample size			Mean weight loss	Sample size			Mean weight loss
	25[c]	50[d]	100		25	50	100	
Onion Storage 1953–54								
Mar. 18	61.3	49.3	32.7	5.4	51.2	24.0	16.3	13.0
Apr. 10	47.7	40.6	26.6	10.7	27.5	20.5	12.1	18.4
May 10	17.4	17.0	10.0	48.6	14.2	13.2	7.9	44.7
Onion Storage 1954–55								
Dec. 20	45.9	33.4	22.5	11.1	25.9	19.5	11.5	18.7
Jan. 22	40.1	29.1	19.7	13.8	21.9	14.9	10.6	21.0
Feb. 19	37.0	27.1	19.3	15.3	20.1	14.0	9.9	22.2
Mar. 26	33.7	22.9	20.3	22.0	18.5	12.8	10.6	26.8
Apr. 23	16.5	10.2	10.6	51.2	14.9	9.8	8.0	45.9
Horticulture Basement 1954–55								
Dec. 21	41.3	25.2	22.1	15.6	21.2	13.3	11.8	23.1
Jan. 21	28.4	27.5	23.3	27.2	15.9	14.9	12.4	31.2
Feb. 18	16.2	13.7	11.2	48.9	11.3	9.4	7.3	44.5
Mar. 26	13.9	9.6	7.4	71.7	10.7	8.1	5.9	58.7

[a]Data analyzed as per cent weight loss.

[b]Percentages transformed to angles by arc sine transformation.

[c]Average CV of four analyses of the four 25-bulb samples, all of which are estimates of the same parameter.

[d]Average CV of the two analyses of the two 50-bulb samples.

pares results from different storages or from the same storage in different years.

Since the coefficient of variation is a ratio of a standard deviation to its mean it would appear helpful to look at these components in interpreting changes in the value of the coefficient of variation. The decrease in the coefficient associated with the longer storage period is due to the increase in mean per cent weight loss, since the standard deviation also tended to rise slightly. However, there was some indication that the standard deviations would again decrease as the mean per cent loss increased beyond 50 per cent. This is what would be expected with data which were binomially distributed, but these data are insufficient to permit the authors to reach any such conclusion. The decrease in the coefficient associated with larger sample size was due entirely to a decrease in the size of the standard deviation since there was also a small decrease in the mean. The decrease in the coefficient arising from transforming to the arc sine is somewhat more complicated. In all cases, the transformations reduced the standard deviation. Due to the character of the transformation, the transformed mean is greater than the percentage mean up to a mean loss of 38.2 per cent. Beyond this the transformed mean is less than the percentage mean. For this reason, the transformation had greater effect at the earlier dates and less on the last date.

Often the plant breeder wishes to use the information given in Table 1 as a guide in deciding how many replications to use. The procedure used here is one described by Neyman and Tokarska (8). The results are necessarily conservative because they are based on the assumption that there are 2 (n—1) degrees of freedom for estimating the standard error of a difference, where n is the number of replications. In a regular storage trial, arranged as a randomized block, there would be several varieties and this error term would be estimated by (r—1) (n—1) degrees of freedom where r is the number of varieties in the trial. Actually this difference causes only a slight dcerease in the size of the difference listed in Table 2. This table gives an estimate of the size of difference expressed as a percentage of the over-all mean that would be required to be detected at the .05 probability level (Type I error) when there are four, six, and eight replications.

Following is an example of how these estimates were derived: $\Delta = \rho\sigma$ where Δ is the estimated difference, ρ is a value found in Table 1 of Neyman and Tokarska's paper, and $\sigma = \sigma_0 \sqrt{2/n'}$ where σ_0 is the coefficient of variation, and n' is the number of replications. In this case, $\rho = 4.28$, with ten degrees of freedom and a chance of non-detection of .01. Also $\sigma = 10\sqrt{2/6} = 5.77$ when the coefficient of variation is 10 per cent. Thus $\Delta = (4.28)\ (5.77) = 25$, which is the figure found in row three and column four in Table 2.

Variability in rotting and sprouting:—The results in Table 3 show that sample size had no effect on either rotting or sprouting. The influence of variety on rotting is not clear cut, but apparently there is a relatively high threshold of random infection which masks any varietal influence up to that point. If conditions are such that there

Table 2.—Size of real differences in percentage of the mean required to be detected at the five per cent level.

CV	Chance of non-detection								
	4 replications			6 replications			8 replications		
Per cent	*.01*	*.05*	*.10*	*.01*	*.05*	*.10*	*.01*	*.05*	*.10*
6	19	16	14	15	12	11	13	10	9
8	26	21	19	20	16	15	17	14	12
10	32	27	24	25	20	18	21	17	15
12	39	32	28	30	25	22	25	21	19
14	45	37	33	35	29	26	29	24	22
16	—	42	38	40	33	29	33	28	25
18	—	48	42	45	37	33	38	31	28
20	—	—	47	49	41	36	42	35	31
22	—	—	—	—	45	40	46	38	34
24	—	—	—	—	49	44	—	42	37
26	—	—	—	—	—	47	—	45	40
28	—	—	—	—	—	—	—	48	43
30	—	—	—	—	—	—	—	—	46

is considerable rotting, then varietal influence plays an important role. The influence of variety on sprouting is large and unquestionable. The magnitude of the variation within varieties does not appear to be very different from one variety to another.

Table 3.—Variability in numbers of rotted and sprouted bulbs expressed in terms of chi square.

Location of storage and date	Rotted bulbs		Sprouted bulbs	
	Among samples χ^2	Among varieties χ^2	Among samples χ^2	Among varieties χ^2
Onion storage 1953–54				
Mar. 18	1.23	6.69	1.22	107.24**
Apr. 10	0.69	8.26*	0.95	613.33**
May 10	0.48	9.00*	3.03	1214.24**
Onion storage 1954–55				
Dec. 20	2.63	3.33	1.75	639.65**
Jan. 22	1.70	2.83	1.21	756.87**
Feb. 19	0.64	2.12	0.17	847.61**
Mar. 26	0.82	2.71	1.79	1339.22**
Apr. 23	0.84	4.92	3.02	2495.02**
Horticulture basement 1954–55				
Dec. 21	0.62	36.37**	1.22	129.49**
Jan. 21	2.04	52.75**	0.40	648.67**
Feb. 18	0.13	61.51**	1.28	1046.23**
Mar. 26	0.29	84.65**	0.94	632.98**

*Significant at P = .05.
**Significant at P = .01.

Test for non-homogeneity of variances:—Chi square values for the non-homogeneity of variances within the four varieties in the several sample sizes and dates studied are given in Table 4. The non-homogeneity was greater in the better storage, where differences in percentage loss were more pronounced. Further, the influence of the transformation in reducing the non-homogeneity was much less on those dates where the mean percentage loss ranged somewhat between 25 and 75 per cent. Such was the case on May 1954, April 1955 in the onion storage, and February 1955 in the horticulture basement. On this latter date, the mean percentage losses were 33.9, 43.2, 46.4, and 72.2.

Table 4.—Chi square for homogeneity of variance.

	Sample size					
	25 bulbs		50 bulbs		100 bulbs	
	per cent[a]	*sin*$^{-1}$[b]	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$
Onion storage 1953–54						
Mar. 18	6.71	2.49	5.25	2.87	5.17	1.99
Apr. 10	4.63	1.73	7.91*	2.54	16.97**	7.82*
May 10	4.40	4.97	3.57	3.29	0.71	2.83
Onion storage 1954–55						
December 20	19.75**	8.49*	8.30	5.63	14.97**	4.98
January 22	15.00**	5.43	11.29*	5.00	10.84**	4.24
February 19	14.35**	5.31	16.30**	9.35*	26.07**	14.43**
March 26	22.72**	10.01*	10.09*	4.12	25.34**	13.54**
April 13	4.82	6.52	4.02	7.14	5.02	3.36
Horticulture basement 1954–55						
December 21	6.33	4.74	2.94	6.35	3.64	3.13
January 21	7.95*	5.93	8.46*	9.59*	19.57**	15.39**
February 18	1.41	2.17	4.03	5.87	0.49	0.71
March 26	6.61	3.97	3.00	3.66	3.33	0.89

[a]Per cent = data analyzed as percentages.
[b]$\sin^{-1}$ = data analyzed as percentages transformed to angles by the arc sine transformation.
*$\chi^2_{.05}$ (3 d.f.) = 7.815
**$\chi^2_{.01}$ (3 d.f.) = 11.341

Test for non-additivity:—Table 5 gives a comparison of F values for the one degree of freedom for non-additivity as computed by Tukey's method (10) for analyses of the percentage and transformed percentage data. The most impressive reduction in non-additivity

Table 5.—F values for non-additivity.

Onion storage 1953–54

Sample size	March 18		April 10		May 10	
No. of bulbs	*per cent*[a]	*sin*$^{-1}$[b]	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$
25	10.26**	6.64*	31.32**	3.32	2.12	4.53
50	0.64	0.12	28.33**	4.25	1.73	2.17
100	40.66**	7.14*	237.24**	27.20**	6.16	8.93**

Onion storage 1954–55

Sample size	December 20		January 22		February 19	
No. of bulbs	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$
25	17.58**	3.74	61.34**	11.94**	58.72**	18.85**
50	3.71	2.51	16.29**	6.70*	27.23**	14.91**
100	40.60**	7.02*	26.74**	3.04	69.64**	17.69**

Sample size	March 26		April 23	
No. of bulbs	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$
25	124.82**	39.09**	18.13**	17.19**
50	87.15**	39.07**	9.82**	8.85**
100	35.79**	13.56**	2.58	1.34

Horticulture basement 1954–55

Sample size	December 21		January 21		February 18		March 26	
No. of bulbs	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$	*per cent*	*sin*$^{-1}$
25	2.23	0.82	7.12*	5.09*	5.17*	6.24*	0.35	0.73
50	0.62	0.03	43.39**	13.77**	30.27**	26.04**	0.98	12.77**
100	6.17*	0.80	77.79**	37.36**	6.54*	3.75	0.26	0.71

*Significant at P = .05
**Significant at P = .01
[a]Per cent = data analyzed as percentages.
[b]$\sin^{-1}$ = data analyzed as percentages transformed to angles by the arc sine transformation.

was on those dates when the mean percentages loss fell outside the range of 30 per cent to 70 per cent. Even so, for many of these dates, the transformed data still had a significant F value for non-additivity.

The binomial portion of the variance:—Table 6 gives the fraction of the total sum of squares attributed to regression on p_i $(1-p_i)$ for

Table 6.—Proportion of the total corrected sum of squares attributable to regression on p_i $(1-p_i)$.

	25 bulb samples		50 bulb samples		100 bulb samples	
	per cent[a]	*sin*$^{-1}$[b]	*per cent*[a]	*sin*$^{-1}$	*per cent*[a]	*sin*$^{-1}$
Onion storage 1953–54	.525**	.005	.530**	.337**	.309	.189
Onion storage 1954–55	.595**	.330**	.432**	.201**	.292*	.143
Horticulture basement 1954–55	.527**	.396**	.617**	.606**	.234	.071

*r significantly different from 0 at the five per cent level.
**r significantly different from 0 at the one per cent level.
[a]Per cent = data analyzed as percentages.
[b]$\sin^{-1}$ = data analyzed as percentages transformed to angles by the arc sine transformation.

the three sample sizes and three locations for both the percentage and transformed percentage data. In all cases the fitting of the regression on p_i $(1-p_i)$ removed a greater proportion of the total corrected sum of squares of the percentage data than of the transformed data. There is no evidence of any location effect. It should be noted that, in general, the differences in the fraction of total sum of squares attributable to regression between the percentage and transformed data are not as large as would be desirable. Thus this transformation had a positive influence in removing the dependence of the variance on the mean per cent loss, but perhaps another transformation could be found which would have a greater influence.

Discussion and Summary

This problem was designed to obtain information on methods of designing onion storage trials and the analysis of the data obtained from them. A storage trail was run two years, and during the last year at two locations, one designed for onion storage and the other the basement of the horticulture building, where the temperatures were high and the humidity low. The experiment was designed as a split plot with size of sample as the main plot and varieties as the subplot. Three sizes of sample, 25 bulbs, 50 bulbs, and 100 bulbs, and four varieties were used. Storage records were taken at monthly intervals during the last part of the storage period.

All analyses were computed both on the basis of percentage weight loss and on the percentage transformed to angles by the arc sine transformation.

Both sets of data were subjected to tests for non-homogeneity of variances and non-additivity, and to the test of the hypothesis that there was no binomial variation.

A table is given which gives the estimated size of a real difference in weight loss expressed as a percentage of the mean that would be required to be detected at the five per cent probability level (Type I error) at three levels of probability of failure to detect any difference (Type II error) when the number of replications is four, six, or eight. The method is outlined for obtaining the same information for different numbers of replications.

It would be to the advantage of the onion breeder to be able to accelerate his storage losses to obtain his storage information earlier. One procedure would be to use less than optimum storage conditions. The data at hand do not recommend such a practice since there exists a large variety-by-location interaction. The effect of changing storages is not one merely of acceleration. Thus it appears necessary to select a storage comparable to those in use by the growers for whom the hybrids are being developed. Another procedure proposed for obtaining storage information earlier, is merely to take the data earlier in the season. Here again, this is complicated by a large variety-by-length-of-storage interaction. Thus the length of storage should be conditioned by the length of time the growers usually need to store their crop.

Size of sample apparently has some effect on storage life both as a main effect and as an interaction with variety, but these effects are small and inconsistent. Any deviations from a common sample size in the range of 25 to 100 bulbs probably would not seriously affect any interpretations made by different breeders, particularly when the same hybrids are tested.

The transformation of the percentage data to the arc sine reduced the non-homogeneity and non-additivity of the variances found in the percentage data. The greatest benefit was derived for those analyses where the mean per cent loss for one or more of the varieties fell outside a range of 30 to 70 per cent. Where there are many deviations from this central range, the data should certainly be transformed, not only to meet more adequately the assumptions of the analysis of variance, but also to increase the efficiency of estimates of differences of the means.

Literature Cited

1. Bliss, C. I. 1937. The analysis of field experimental data expressed in percentages (in Russian with English summary). *Zaschita Rastenii (Plant Protection)* 12:67–77.
2. ————. 1938. The transformation of percentages for use in the analysis of variance. *Ohio Jour. Sci.* 38:9–12.
3. Clark, A., and W. H. Leonard. 1939. The analysis of variance with special reference to data expressed as percentages. *Jour. Amer. Soc. Agron.,* 31:55–66.
4. Cochran, W. G. 1938. Some difficulties in the statistical analysis of replicated experiments. *Empire Jour. Exp. Agr.* 6:157–175.
5. ————. 1940. The analysis of variance when experimental errors follow the Poisson or Binomial laws. *Ann. Math.* Statis. 11: 335–347.
6. ————. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics* 3:22–38.
7. Eisenhart, Churchill. 1947. The assumptions underlying the analysis of variance. *Biometrics* 3:1–21.
8. Neyman, J., and B. Tokarska. 1936. Errors of the second kind in testing student's hypothesis. *Jour. Amer. Statis. Assoc.* 31:318–326.
9. Snedecor, G. W. 1946. Statistical methods. 4th ed. Ames, Iowa, Iowa State College Press.
10. Tukey, J. W. 1949. One degree of freedom for non-additivity. *Biometrics* 5:232–242.
11. Woodman, R. M., and H. R. Barnell. 1937. The connexion between the keeping qualities of commercial varieties of onions and the rates of water loss during storage. *Ann. Appl. Biol.* 24:219–235.