Research Report on

Washington Winter Wheat Objective Yield Estimates

by

Fred B. Warren

Research and Development Branch

Standards and Research Division

Statistical Reporting Service

October 1969

Some Effects of Sample Unit Location Procedures on Washington Winter Wheat Objective Yield Estimates

General:

The final objective pre-harvest estimates of winter wheat yields for the State of Washington were from 9.7 to 10.3 bushels higher than the Board estimates in three (1965, 1967, and 1968) of the first four years that objective yield estimates were made for that state. In the fourth year, 1966, the difference was only 0.2 bushels. For the last 2 years, the differences were considerably larger than could be explained by sampling error.

Table 1.--Objective and Board final estimates of winter wheat yields, bushels per acre, Washington, 1965-68

Year	Board final	Objective estimates		Difference
		Yield	Sampling error	
1965	41.5	51.2	4.9	9.7
1966	40.5	40.7	2.4	0.2
1967	42.0	51.7	2.4	9.7
1968	40.0	50.3	2.2	10.3

Background:

One theory offered to explain the differences given above involves the procedure used to locate sample units in sample fields. The location procedure required the enumerators to locate the units by walking a randomly selected number of paces, along the edge and into the field, from the most accessible corner. The unit location tables giving the random numbers was designed so that units would fall randomly in a quarter of a 40 acre field, assuming an average 36 inch pace. (In practice, most enumerators would have a shorter pace so that the area which could be sampled would be somewhat less than ten acres). In Washington, particularly in the Palouse area, a number of fields are much larger than 40 acres. Some of these include steep hills and the road pattern in these areas generally follows the valleys. Hence, most fields would be approached from the lower elevations. Sample units in these fields, using the most accessible corner technique, would be located on the (presumably) more productive lower slopes. Under these conditions, the plot selection procedure could be responsible for much of the differences observed between the objective and Board estimates of yield.

This study was undertaken to evaluate the validity of this theory using information available for 1968 from various sources. The sample fields selected in the 1968 winter wheat objective yield survey were classified by such factors as size of field, amount of change of elevation within the fields, location of the starting corner with respect to high and low elevations in the fields, and a comparison of farmer reported yield for the entire field with the sample estimates from objective yield plots.

Distribution of Samples by Fields, by Segments, and by Size of Field

The 1968 winter wheat objective sample consisted of 130 samples selected from the December 1967 and June 1968 Enumerative Surveys. These 130 samples were located in 100 fields and 88 area segments. In 85 of the selected fields one sample was assigned. Two samples were assigned to 10 of the sample fields. Five other fields were assigned 4, 5, 5, 5, and 6 samples respectively.

Four segments were assigned 4 or more samples. One segment (2233) was assigned 16 samples, 12 percent of the state total. This is an excessive number of samples for one segment for effective sampling even though the assignment was unbiased. The fact that 12 of the 16 samples assigned to segment 2233 were drawn from the December Enumerative Survey indicates the sampling procedure for the December Enumerative Survey tracts might be improved if a similar situation exists elsewhere.

In Washington, 23 of the 130 samples were located in fields that were less than 80 acres in size. Fields in the 81 to 320 acre size category contained 83 samples. Six samples were in fields larger than one section (640 acres) in size. See Table 2 for the size distribution of fields. Hence, the location procedure used for laying out sample units would, in theory, reach the center of very few fields.

Distribution of Samples by Variation in Elevation Within the Field

Geological survey contour maps, scaled of one inch to the mile (15 minute quadrangles) or 2 5/8 inches to the mile (7 1/2 minute quadrangles), were available for 62 of the segments having winter wheat samples. Boundaries for these segments and the 92 sample fields were transferred to the geological survey contour maps from contact prints. The amount of change in elevation in the individual fields was determined by inspection of the contour lines on these maps.

The other 38 samples, for which the geological survey contour maps were not available, were classified as being in one of four different categories on the basis of terrain features observed from contact prints. The four categories were (1) moderately flat, (2) rolling, (3) moderately rough and (4) rough. The intent was to classify these sample fields in broad categories by variation in elevation as follows:

Variation in Elevation	Classification of Terrain
50 feet or less	moderately level
51 to 200	rolling
201 to 400	moderately rough
401 or more	rough

Table 2.--Distribution of winter wheat samples by size of field and change in elevation, Washington, 1968

Change in elevation	Size of field in acres					Total
	80 or less	81 to 160	161 to 320	321 to 640	More than 640 1/	
Feet	Number	Number	Number	Number	Number	Number
20 or less	4	5	2			11
21 to 50	4	5	5	1		15
51 to 100	4	2	4			10
101 to 150	5	4	13	1		23
151 to 200		3	6	1		10
201 to 300	2	1	5	3	1	12
301 to 400		1	1		2	4
401 to 500				1		1
More than 500				5 2/	1	6
Sub-total	19	21	36	12	4	92
Elevation unknown	4	15	11	6	2	38
Total	23	36	47	18	6	130

1/ Individual fields in this size group contained, 777, 1430, 844, 749, 680, and 660 acres respectively.

2/ Only one field in this group (with 5 samples) had a total change of elevation of 550 feet.

Table 3.--Distribution of samples by size of field and classification of terrain, Washington, 1968

Terrain classification	Acres in field					Total
	80 or less	81 to 160	161 to 320	321 to 640	More than 640	
	Number	Number	Number	Number	Number	Number
Moderately flat	8	14	7	2	--	31
Rolling	10	11	25	2	--	48
Moderately Rough	5	10	14	7	2	38
Rough	--	--	--	6	3	9
Total	23	35 1/	46 1/	17 1/	5 1/	126

1/ One sample not classified since contact prints were not available.

Table 3 gives the distribution of sample fields by size and terrain classification. Almost a fourth of the samples were located in fields which were classified as being moderately flat (variation in elevation of less than 50 feet). It seems unlikely that the productivity of the sample areas (those closer to the border at the most accessible corner) of these fields would be much different from the productivity of the interior of such fields. Another 48 samples, 38 percent of the total, were located in fields which were classified as "rolling" (variation in elevation of 51 to 200 feet). There is a definite potential for "location bias" both in these fields and in the 47 samples assigned to fields with even greater variation in elevation (201 to 600 feet).

Portion of Field Located Below the Starting Corner

The probable starting corner was determined by examination of the county road maps, contact prints, and geological survey maps. Areas of the field between the contour lines were planimetered to determine the proportion of the field located below the starting corner for 92 samples. For 34 additional samples, the approximate proportion of the field below the level of the sampled area was determined by inspection of 1/16000 scale aerial photographs of the sample fields. Data from this analysis is summarized in Tables 4, 5, and 6.

Table 4.--Number of wheat samples by proportion of field below starting corner and by classification of terrain, Washington, 1968

Proportion of field below starting corner	Classification of Terrain				
	Moderately level	Rolling	Moderately rough	Rough	Total
Percent	Number	Number	Number	Number	Number
0- 5	2	16	15	1	34
6-15	1	4	4	2	11
16-25		6	3		9
26-35		3	2		5
36-45		7	3	1	11
46-55	16	3	7	3	29
56-65	1	2			3
66-75	3	1	1		5
76-85		3	2		5
86-95	5	3	1	2	11
96-100	3				3
All	31	48	38	9	126

There was a definite tendency for the sample starting corners to be located in the lower regions of fields classified as rolling, moderately rough or rough. About one third of the starting sample corners in these classifications were located in the lower five percent of the field. Almost one half were found in the lower fourth of these fields.

The material in Tables 5 and 6 further illustrate the tendency of starting sample corners to have been located in lower portions of those fields classified as being at least "rolling" or having a maximum change of elevation within the field of at least 50 feet. Aside from the tendency of larger fields to have a greater total change in elevation within the field, there appears to be no appreciable relationship between field sizes and the relative elevation of the starting corner.

Table 5.--Average proportion (p) of wheat objective yield sample fields below the leval of sampled areas, by size of field and by amount of change in elevation within the field (92 samples for which geological survey maps were available), Washington, 1968

Change in elevation (feet)	Size of field (Acres)										Total	
	80 or less		81 to 160		161 to 320		321 to 640		More than 640 1/			
	n	p	n	p	n	p	n	p	n	p	n	p
20 or less	4	.42	5	.55	2	.58	-	---	-	---	11	.51
21 to 50	4	.51	5	.90	5	.60	1	.73	-	---	15	.68
51 to 100	4	.24	2	.06	4	.26	-	---	-	---	10	.22
101 to 150	5	.27	4	.30	13	.31	1	.60	-	---	23	.32
151 to 200	-	---	3	.00	6	.35	1	.59	-	---	10	.27
201 to 300	2	.50	1	.06	5	.14	3	.15	1	.05	12	.19
301 to 400	-	---	1	.00	1	.00	-	---	2	.34	4	.17
More than 400	-	---	-	---	-	---	1	.62	-	---	1	.54
Total	19	.37	21	.41	36	.33	7	.54	3	.24	92	.38

Table 6.--Average proportion of wheat objective yield sample fields below the level of sampled areas, by size of field and by type of terrain (126 samples), Washington, 1968

Type of terrain	Size of field (Acres)										Total	
	80 or less		81 to 160		161 to 320		321 to 640		More than 640			
	n	p	n	p	n	p	n	p	n	p	n	p
Moderately level	8	.46	13	.67	8	.58	2	.62	---------		31	.59
Rolling	10	.23	11	.25	25	.33	2	.60	---------		48	.30
Moderately rough	5	.40	9	.35	14	.21	7	.12	3	.24	38	.25
Rough	------		------		------		6	.62	3	.07	9	.34
Total	23	.35	33	.44	47	.34	17	.41	6	.16	126	.36

Comparison of Final Objective Yield Estimates with Farmers Reported Yields

The average yields reported for a sub-sample of fields by farm operators as part of the Form D post-harvest interview was compared with the final objective yield estimates for these same fields. These were not expected to agree exactly since the objective yield estimates from sample plots are not designed to provide field estimates. However, samples showing large differences between the two yield estimates were examined to determine if the location of the starting corner could have been a contributing factor.

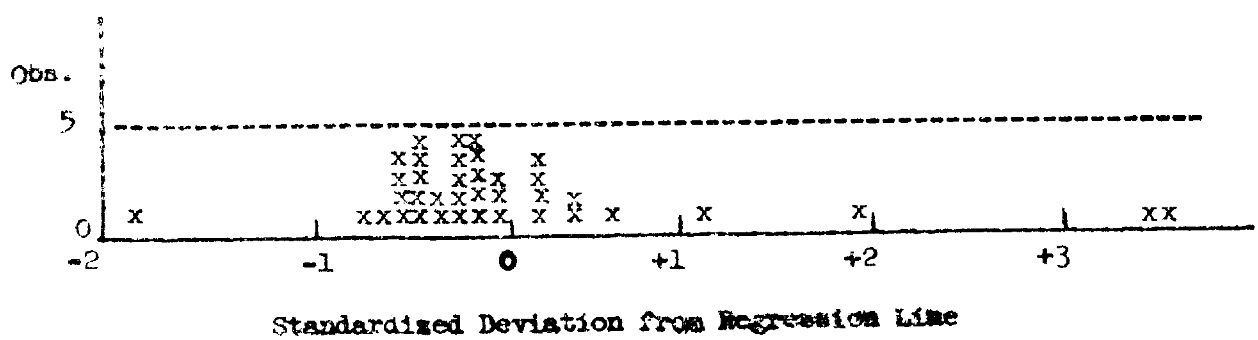
Comparable yield estimates were available for 38 samples. The average Form D yield was 5.6 bushels below the average of the objective yield estimates for the same samples. A test of significance for consistent difference resulted in a t value of 1.89, approaching, but not quite significant at the five percent level of probability.

A regression analysis of the Form D yields upon the comparable Form B objective estimates yielded the regression equation $Y = 9.9881 + .9008X$ where X is the reported yield for sampled field, and Y is an estimate of what the objective yield estimate would be under conditions of perfect correlation. The actual correlation ($r = .68$) was significantly large at the one percent level. The computed intercept and regression coefficient then would reflect a general tendency for the farmers reported yields to be lower than the objective yield estimates for the same fields. This tendency would decrease as the two yield levels increase, vanishing at a yield level of about 100 bushels.

In linear regression, the quantity $S^2_{y.x}$ describes the amount of variation of the actual Y values about the computed regression line. For a given value of X, X_i, $S^2_{y.x_i}$ is computed as $S^2_{y.x}\left[1 + \frac{1}{n} + \frac{X_i - \bar{X}}{S^2_x}\right]$, where n, $\bar{X}$, S_x^2, and $S^2_{y.x}$ relate to the data from which the parameters a and b of the regression equation were computed.

Further, the quantities $\sqrt{\frac{\hat{Y} - Y}{S^2_{y.x}}}$, the standardized deviation from the regression line, should be distributed normally with unit variance and mean zero. The actual standardized deviations plotted in Figure 1 show that they have a median value of -0.25 bushels and that the clustering about the median is much tighter than would be expected if they were distributed normally.

Figure 1.--Standardized deviations of actual objective yield estimates from the Regression line



If the distribution of these deviations was truly normal, we would expect that 16 (about 42 percent) of the observed standardized deviations would have an absolute value of at least 0.8. For 38 paired observations, the probability that only five standardized deviations would exceed a value of 0.8 is less than 1 percent. The fact that four of these five deviations are considerably larger than 0.8 implies that they were taken from different populations. This could happen if there was a large amount of variation in yield within the field and the area sampled for the objective estimate was in an extreme area.

The two most extreme values came from samples 14 and 35, moderately rough or rough fields, where the sample area was located in the lower quarter of the field. The next two most extreme values (one positive and one negative) were from moderately level fields where the sample was taken in the upper fourth of the field. The hypothesis suggested by this analysis is that extreme differences in yields can be associated with plots located in the lower portion of rough terrain type fields.

Table 7.--Characteristics of samples with large deviations from the computed regression line, Washington, 1968

Sample number	Farmer reported yield (X_i)	Objective yield estimate (Y_i)	$\hat{Y}_i = a + bX_i$	$\frac{Y_i - \hat{Y}_i}{S_{y.x}}$	Size of field	Type of terrain	Proportion of field below starting corner
	Bushels	Bushels	Bushels	Bushels	Acres		Percent
14	63.3	129.8	67.0	3.23	174.6	Moderately rough	25
35	46.0	115.4	51.4	3.33	680	Rough	10
80	61.1	101.5	65.0	1.88	30	Moderately level	90
113	80.0	44.1	82.1	-1.89	155	Moderately level	75
125	40.0	66.0	46.0	1.04	112.2	Unknown	--

Comparison of Objective Yield Averages with Board Estimates by Counties

If the sample layout does cause sample units to be located in the more productive areas of the fields with variable elevations, then we might expect the average of the objective yield sample estimates to be considerably higher than the true county average where a relatively large proportion of the sample fields would be classified as "rough" or "moderately rough". Using the Board's 1968 county estimates as a yardstick, (Table 8), we find the greatest differences between the objective yield and Board estimated yields for major counties occur in Benton, Franklin, and Lincoln Counties. Five of the nine samples in Benton

Table 8.--Comparison of objective yield estimates and sample allocation with board estimates by countries, 1968

County	Acres in County	Objective yield samples			Indicated Yields		
		Total	Classified as Rough	Classified as Moderately Rough	Objective	Board	Ratio
	(000)	(Number)	(Number)	(Number)	(bu.)	(bu.)	(bu.)
Whitman	452.6	20	--	3	52.7	48.0	1.10
Adams	398.4	17	--	3	37.2	38.0	.98
Lincoln	388.4	19	2	8	58.2	41.0	1.42
Walla Walla	242.8	6	1	1	35.7	38.0	.94
Grant	212.3	9	--	2	59.8	48.5	1.23
Douglas	204.4	9	--	6	38.8	30.4	1.28
Benton	143.0	0	5	--	34.1	16.0	2.13
Franklin	132.8	20	1	--	62.1	38.5	1.61
Spokane	121.0	5	--	1	39.9	48.0	.83
Columbiana	90.6	4	--	3	58.2	46.0	1.27
Garfield	85.4	4	--	4	53.5	44.0	1.22
Klickatat	65.2	3	--	--	40.9	32.0	1.28
Yakima	41.4	3	--	1	100.2	42.0	2.38
Asotin	32.2	1	--	--	50.9	26.5	1.92
Stevens	19.2	1	--	--	97.2	43.0	2.26
Other	25.3	0	--	--	----	*	----
Total	2655.0	130	8	32	50.5	40.0	1.26

county were in a single field which was classified as rough, with an actual change in elevation in excess of 500 feet. It so happened that all but one of the starting corners of this field would have been located on comparatively level ridge tops. The objective yield estimate for Franklin county was influenced by that sixteen samples assigned to segment 2233. This segment is in an area which was designated as rangeland in the sampling frame but where the land use has now changed to irrigated wheat. This segment has very high yields and accounts for possibly 2 to 3 bushels of the difference between the objective yield estimate for the State and the Board estimate. Ten of the nineteen samples in Lincoln county were in fields classified as being at least moderately rough. None of the samples were located in the upper half of the field. The objective yield average (53.8 bushels) of these 10 samples was only 4.2 bushels less than the 58.0 average for all samples for the county.

Summary

There has been a definite tendency for samples in fields with steep slopes to be located in either the lower or upper, generally lower, portions of the fields. To the extent that the lower portions of such fields are more fertile, retain more moisture, and are less subject to wind damage, this could lead to:

(1) Overestimation of the average number of heads per acre for the field.

(2) Overestimation of the average weight of grain per head for the field.

To measure the effect, if any, of these factors would require a special validation type survey, making final pre-harvest and post-harvest observations in two sets of sample units in a number of fields. One set would be located using the sample allocation used in Washington prior to 1969 (and still used by all other wheat states). The second set of units would be instituted for Washington starting with the 1969 crop season. Comparison of the paired sets of yield components obtained using the two methods of sample location would indicate the effect of the previous procedure.