CHAPTER VII-TN 15: DEPTH OF SNOWFALL AND ITS RELATION TO DISTANCE FROM THE CONTINENTAL DIVIDE AND ELEVATION ABOVE SEA LEVEL

By H.K. Cheung and J. Beaman

ABSTRACT

The *prima facie* objective of this paper is to present the results of determining the parameters of an equation to be used to predict snowfall in terms of the two variables, elevation and distance from the continental divide. It is actually provided because it was found that readers of CORDS documents needed a simple and easily understood application of the regression, "analysis of variance", technique that is important in TN 12, 6, 29, 4 etc. A simple application to a problem in which results are easily visualized was deemed to be using the method to study how snowfall related to elevation and distance form the divide.

OBJECTIVE

This paper presents the results of predicting snowfall in terms of the two variables, elevation and distance from the continental divide. It will prove useful to the reader who is not familiar with the regression in which a dependent variable is explained by effects associated with categories of independent variables (with ranges of elevation or distance). The simple model being spatially interpreted can yield an intuitive grasp how the analysis works. This leads to an understanding of the kind of analysis that can be carried out when one has variables for which values are categories.

APPROACH TO ANALYSIS

The data given in Table 1 were available for analysis. These data provide information that may suggest a distance or elevation relation of some significance in searching for potential ski areas on the basis of available information on elevation and location of the divide. Elevation and distance from the divide could also be treated as continuous (interval) variables.

TABLE 1: SNOWFALL DATA FOR THE CANADIAN ROCKIES (IN THE AREAS INDICATED)

Snow Station	Elevation (feet)	West Distance from Divide (mi.)	Average Annual Snowfall (19 yr average)
Canmore	4350	24	62
Kanaskia Cabin	5500	8	134
Kananaskis Lookout	6800	10	151
Pocatterra Creek	6720	12	123
Highwood Pass	7250	6	173
Mud Lake	6220	8	180
Evans—Th. Creek	4950	20	103
Whiteman Pass	5600	12	85
Elbow Lake	7150	12	175
Exshaw	4200	21	76
Kanaskis Station	4560		102
Anthracite	4550	20	64
Banff Town	4582	17	79
L. Louise	5032	5	193
Sunshine (3 year average)	7200	0	308
Norquay (3 year average)	6120	17	100

There was no reason to expect a linear relationship (or any particular relationship, for that matter) between the variables considered. Given the limited data available, it was decided to use an "analysis of variance model", today generally called a regression model with dummy variables (see ORCOL 1977), to determine effects that are described later. Use of this model allows data to be grouped into categories. This makes it possible to avoid the assumption that snowfall is a function of distance from the divide and elevation of the form:

$$\text{Snowfall} = U + C_1(\text{elevation}) + C_2(\text{distance from divide})$$

WHERE U, C_1 and C_2 are constants that would be determined by regression.

The categories used in grouping data were selected to correspond with "natural" groupings in the data. This resulted in enough responses in categories so that the parameters to be estimated (could) be estimated. The categories used are given in Table 2.

TABLE 2: CATEGORIES USED IN THE ANALYSIS OF VARIANCE

ELEVATION		DISTANCE FROM THE DIVIDE	
Category	Elevation in feet above sea level	Category	Distance from divide in miles
1	Up to 4999	1	0 - 6
2	5000 - 5999	2	7 - 12
3	6000 - 6499	3	13 - 19
4	6500+	4	20+

TABLE 3: ESTIMATED PARAMETER VALUES

U=General Snowfall Level = 130.95 inches

ELEVATION EFFECTS			DISTANCE EFFECTS		
Category	Categories(Elevation in feet)	Effect	Category	(Miles from divide)	Effect
1	Up to 4999	9.33	1	0 -6	106.61
2	5000-5999	-41.63	2	7 - 12	18.72
3	6000-6499	30.33	3	13 - 19	-61.28
4	6500+	0.97	4	20 +	-64.03

The model for which estimation was carried out can be interpreted as follows:

Total snow fall for elevation category X and distance form the divide category Y	=	Effect for elevation category X	+	Effect for distance form divide category Y	+ Error

RESULTS OF ESTIMATION

Many ways exist to estimate the "effects" that the equation above indicates must be estimated. They are outside the scope of this paper. For discussion of methods using standard regression programs, the dummy variables method, see ORCOL 1977. When analysis of variance of snowfall data was carried out, a fairly good explanation was achieved. A usual criterion for goodness of explanation is the ratio R^2 which is the explained sum of squares to the total sum of squares. An R^2 of .79 was achieved. Estimated values of coefficients for which results were obtained are given in Table 3. To show how well the predictions correspond with observations, Table 4 presents both observed and predicted results.

Table 4 actually has several functions. By looking at it, one can see the effects that were added together to get a given observation. Given a hypothetical situation of the snowfall at 7,500 feet at 32 miles from the divide, the estimate of snowfall is obtained just as it is in Table 4 for a known site. Also, from the estimated parameter values in Table 3 it is seen that the combination of 6000-6499 feet of elevation and 0-6 miles of distance from the divide will have the largest positive effect on snowfall. Whether this is reasonable requires a thorough knowledge of the geography of the study area.

Snowfall = general snowfall + (6,5000 + ft elevation effect) + (20+ miles distance effect)

There is the possibility that the residuals shown in Table 4 show some kind of systematic error. There could be a distance-elevation interaction meaning that combinations of distance from the divide and elevation may be conducive to high or low snowfall in a way that is not expressed by simply adding an elevation effect and a distance effect. Patterns such as these do not seem apparent in the data.

It is also possible that large differences between observations and predictions that are seen in Table 4 may be related to a third variable, which was not considered in this analysis. Obvious variables to consider are orientation and steepness of the slope on which the measurement was taken. Or the residuals may indicate something about the nature of the divide near the area being considered. Large negative values of residuals may indicate one kind of divide terrain while positive residuals may indicate quite different geographic characteristic of the nearby divide. Such factors are likely to be picked out by somebody with a "feel" for the areas but not likely to be found in a purely *ad hoc* statistical analysis.

DISCUSSION OF THE RESULTS

When examining Table 4 one sees that the range of residuals varies from -66 to 70 with an average error of 14.2 percent. It is interesting to note that among the residuals, the two largest (-66 and 70) are associated with observation units that are at short distances from the divide and have high elevations. Reasons for this large discrepancy are not immediately apparent from the data. Some variables that are of importance in explaining snowfall (and that were excluded from the analysis) may account for these large residuals. Nevertheless, this model as it is may yield poor prediction of snowfall for an observation unit at a high elevation that is also at a short distance from the divide. On the other hand, it seems clear that it may give relatively good predictions for those observation units for which other combinations of elevation and distance from the divide exist.

TABLE 4: OBSERVED AND PREDICTED VALUES FOR THE OBSERVATION UNITS

Unit	Obs. Snowfall	Predicted Snowfall (inches)	Residual	Elevation Category	Distance from Divide Category.
01	62	76 = 131 + 9 - 64	- 14	1	4
02	134	108 = 131 - 42 + 19	26	2	2
03	151	151 = 131 + 1 + 19	0	4	2
04	123	151 = 131 + 1 + 19	- 28	4	2
05	173	239 = 131 + 1 +107	- 66	4	1
06	180	180 = 131 + 30 + 19	0	3	2
07	130	76 = 131 + 90 - 64	27	1	4
08	85	180 = 131 - 42 + 19	- 23	2	2
09	175	151 = 131 + 1 + 19	24	4	2
10	76	76 = 131 + 9 - 64	0	1	4

11	64	76 = 131 + 9 - 64	- 12	1	4
12	79	79 = 131 + 9 - 61	0	1	3
13	193	193 = 131 - 42 +106 = 195	- 2	2	1
14	308	238 = 131 + 1 +106 = 238	70	4	1
15	100	100 = 131 + 30 - 61 = 100	0	3	3

The use of average annual snowfall as the dependent variable may present some problems, since large extreme values of snowfall were observed. Since the mean snowfall is affected appreciably by large extreme values, the median might have been a better dependent variable. The use of time series data offers the advantage that the situations being modelled are more truly reflected in that variations that occur over time are open to explanation. On the other hand, the use of time series data requires the specification of how dependent variable values relate to time. A valid relation may not be easy to arrive at.

When analysis is carried out on more data, it will be possible to correct for different variances on observations and intercorrelations between observations. The present analysis simply ignores the fact that observations at different stations are almost certainly correlated. Also it is reasonable to think that there will be more variation (absolute) in amount of snow in high snowfall areas than in low snowfall areas. Though the lack of knowledge about these factors does not cause a bias in the results presented, it lowers the efficiency of the estimates, meaning that better predictions could be made.

It may have occurred to some readers that the parameter values in Table 3 are difficult to accept in a practical sense. For instance, the difference between elevations 5999 feet and 6000 feet is only one foot and yet the difference in the effect of snowfall is almost 72 inches. The difficulty clearly is one of interpretation. The reader is reminded that the explanatory variables used are not continuous but are categorical or classificatory. As such, the parameters presented must be interpreted accordingly: the difference in effect on snowfall between elevation category 2 and category 3 is almost 72 inches. One could associate parameter values with a category mean and then draw a curve through the parameter values. The "appropriate" parameter for an elevation could be read from the curve and this estimate would not suffer from the problem just noted.

CONCLUSION

The model derived suggests a reasonable means of predicting snowfall for a weather station in terms of its elevation measured in feet and its distance from the divide measured in miles. There is no assumption that depth of snowfall only varies with the two variables considered; and the results suggest that it does not. It is only assumed that the effects on snowfall due to the explanatory variables can be separated so that the overall effect is the sum of the individual effects. Interaction between the explanatory variables used was not hypothesized, since such a relationship was not apparent from the data provided. Exclusion of explanatory variables that are significant in predicting snowfall (such as the nature of the terrain) may have caused the two large residuals mentioned before. Of course, the adequacy of the model to predict can only be determined after it has been applied to additional sets of data.