From the New York University and the Institute of Nutrition of Central America and Panama

An application of the analysis of covariance to dietetic Studies in Acatenango, Guatemala

By CH. ROBERTS and M. FLORES

With 6 tables

(Received July 22, 1968)

Introduction

Much of the variation of caloric intake of an individual can be explained by his caloric requirement. Taking an extreme example, it is obvious that if his requirement is 2000 calories daily, one does not find this individual with a 20,000 calories per day intake. Furthermore, he is not on a 200 calories per day diet or he would have long since died of starvation.

By removing the effect of requirement, it may become possible to study some aspects contributing to caloric intake, which may be very important to the dietitian. For instance, whether people have higher caloric intakes on the weekend. Because a person is essentially eating his requirement, it may not be possible to see an effect of day of the week on intake, although it may exist, due to the large size of the numbers involved. If the effect of requirement is removed, what is left is much smaller so that in comparisons from day-to-day, a small change might point out an importance previously ignored.

The idea of removing the effect of requirement can be extended to other nourishment factors, such as vitamins and animal protein, although the extension will not be attempted here. If the requirement of a certain vitamin is unknown, it could be assumed proportional to the requirement in calories. In other cases, it may be assumed that the requirement is proportional to the total weight of the individual, or to his age, as a crude approximation. Therefore, although the examples considered will deal only with calories, the theory can applied to other nutritional factors, whether or not they have well-known or tabulated requirements.

The data

Forty families were selected at random from the town of Acatenango, and each family was studied exactly one day. The data is presented in table 1, where it can be seen that 6 different families were studied on each of 5 days, 8 families studied one day, and only 2 families studied on the last day.

Ten other families were selected at random from the same town and studied for an entire week. The data observed is presented in table 2.

The nutritionists who collected the data lived in Acatenango during the period of data collection so that they could easily visit homes daily before breakfast had been prepared and in the evening for the last meal. They carried scales to weigh the cooked and uncooked products and made liquid measurements when applicable. Samples of local fruits and vegetables were obtained for comparison with established standards.

Table 1. Consumption and Requirements of Calories for the One-day Method (Model I)

No. of Case	Day of Study: First[1]) C[3])	First[1]) R[4])	Second C	Second R	Third C	Third R	Fourth C	Fourth R	Fifth C	Fifth R	Sixth C	Sixth R	Seventh[2]) C	Seventh[2]) R
1	9.577	8.800	16.752	17.589	15.008	16.564	13.268	12.500	12.754	8.220	23.229	17.070	7.442	4.748
2	5.847	10.200	22.682	21.409	15.413	22.220	8.008	7.100	14.960	13.600	15.108	9.650	12.229	12.592
3	9.615	11.630	24.311	22.340	6.707	5.620	4.833	6.970	27.991	17.862	16.017	14.762		
4	15.746	13.380	18.922	19.430	33.788	30.155	29.963	26.242	16.554	13.240	26.702	20.930		
5	10.356	10.350	12.972	12.520	19.274	21.510	17.880	13.322	30.455	25.275	20.284	15.500		
6	10.009	9.770	7.632	6.910	32.902	16.498	12.052	10.750	8.915	11.620	21.374	13.900		
7											6.182	9.600		
8											6.804	8.514		

[1]) Thursday, December 2, 1965.
[2]) Wednesday, December 8, 1965.
[3]) Thousands of Calories Consumed.
[4]) Thousands of Calories Required.

Table 2. Consumption and Requirements of Calories for the Seven-day Method (Model II)

Family Number	Calories Consumed[1])	Requirement[2])
1	18.744	15.898
2	18.035	12.961
3	10.546	10.387
4	10.186	9.067
5	4.753	7.188
6	17.716	16.791
7	15.109	13.070
8	11.316	10.036
9	11.246	8.340
10	10.387	7.410

[1]) Average daily consumption in thousands of calories for the entire family during the week of December 2–8, 1965.

[2]) Average daily requirement in thousands of calories for the entire family during the week stated.

It will be shown in the following section about the applicable mathematical models that it was indeed unfortunate that the intakes of each family were not recorded on a daily basis for the seven-day method. For instance, if a family cooked 5 pounds of black beans on a Monday and ate them over a period of three days, it was recorded only that during this three-day period the family consumed the 5 pounds of beans. This resulted in the fact that in the analysis it is not possible to see the effect of family or of day of the week.

To dry to salvage the data for the seven-day method, the amount that was consumed by the family on a daily basis was estimated. If a quantity of beans was eaten over a period of two meals or three meals, one-half or one-third was assigned as having been eaten in each meal, respectively. This is clearly an approximation since probably more is eaten in the first meal, less in the second, and so on. Yet it is hoped that on the average the results will be relatively accurate, and, in any case, it provides us with an interesting example. The corrected or approximated data is to be found in table 3. The resulting data now lacks independence but the tests to be made should still be reliable. If the data had been recorded on a day-to-day basis, this difficulty could have been avoided.

The requirement in calories is based upon (1) and takes into account the sex, age, weight, pregnancy and/or lactating, and the average temperature (20 °C in Acatenango). The computation of the daily caloric requirement took into account the attendance of each person at each meal. If a person missed a meal, one-third of his caloric requirement for the day was removed from the family daily requirement, which is believed to be a good approximation in Acatenango, where the died is essentially the same for all meals, and if a person knew that he was going ro be away at lunch, he would eat more at breakfast.

Table 3. Consumption and Requirements of Calories for the Corrected Seven-day Method (Model III)

Family Number	Day of Study: First[1] C[3]	First[1] R[4]	Second C	Second R	Third C	Third R	Fourth C	Fourth R	Fifth C	Fifth R	Sixth C	Sixth R	Seventh[2] C	Seventh[2] R
1	17.965	17.940	23.702	17.940	21.409	17.940	5.381	6.893	17.937	17.433	20.009	17.904	24.714	15.200
2	22.816	13.570	12.681	13.570	18.119	12.060	20.693	13.067	20.762	11.323	16.476	13.570	14.695	13.570
3	8.179	8.920	8.007	8.920	10.268	9.710	12.069	11.290	12.700	11.290	10.508	11.290	12.094	11.290
4	11.543	8.427	11.160	8.427	10.627	10.220	11.111	10.220	7.778	10.220	10.508	8.427	8.578	7.530
5	6.072	7.170	4.715	7.170	3.813	7.170	3.795	7.170	5.331	7.170	5.040	7.170	4.506	7.294
6	22.058	16.480	21.482	16.480	18.141	16.480	19.203	17.870	15.520	16.480	14.539	16.480	13.066	17.270
7	14.238	13.070	14.451	13.070	14.907	13.070	16.220	13.070	14.884	13.070	13.168	13.070	17.892	13.070
8	14.943	10.007	12.164	9.110	11.249	11.800	11.324	11.800	13.576	11.800	2.829	3.933	13.128	11.800
9	13.432	8.340	12.159	8.340	10.180	8.340	9.984	8.340	10.087	8.340	10.911	8.340	11.968	8.340
10	10.940	7.410	8.462	7.410	10.782	7.410	11.458	7.410	11.577	7.410	10.619	7.410	8.871	7.410

[1]) Thursday, December 2, 1965.
[2]) Wednesday, December 8, 1965.
[3]) Thousands of calories consumed.
[4]) Thousands of calories required.

The mathematical models

The three models that will be considered here are studied mathematically in (2).

In the first study mentioned, in which the family was studied only one day, because there were no replications (repetitions) of the experiment, it is impossible to see a family effect. Since the family was considered as a whole, it is not possible to analyze the effect of the single individual. Therefore, the appropriate model is

$$C_{js} = a + d_s + v\,(R_{js} - \bar{R}) + E_{js} \qquad \text{[I]}$$

where C_{js} is the number of calories consumed by the family j, on day s.

The quantity R_{js} is the requirement in calories of the family j, on day s, and $\bar{R}$ is the average of the requirements, which has been averaged over family and day. The E_{js} is the error involved which is assumed to be normally distributed and independant of j and s. The number a is a fixed effect and d_s is the effect of day of the week. The quantity v is a type of velocity or the rate at which calories are consumed, in comparison with the required consumption of calories.

In the second study mentioned, in which 10 families were studied exactly one week, because exact intakes were not recorded daily, the model is only a simple regression given by

$$C_j = a + v\,(R_j - \bar{R}) + E_j\,. \qquad \text{[II]}$$

In this example, although seven times as much work was done per family as in Model I since each family was studied en entire week, it can be seen that Model II is less general than Model I because it lacks the day effect.

After the estimation of daily intakes is made for the families which were studied an entire week, the corresponding model is

$$C_{js} = a + b_j + d_s + v\,(R_{js} - \bar{R}) + E_{js} \qquad \text{[III]}$$

where b_j is the effect of family. This model considers the effect of day of the week and the effect of the family when the effect of requirement has been removed. Model III is the most general model that could be considered for use with this data.

Statistical tests and estimation for the one-day method (Model I)

We wish to test the significance of day effect, and estimate v and the variance of E_{js}. The appropriate test is given in table 4. Thus, it is seen that the day effects are not significantly different at the 5% level ($F = .68$). We calculate $v^* = 1.13$, an estimate of v, and var $(E_{js})^* = 17.84015$, an estimate of the variance of each E_{js}. A further test shows that v^* is not significantly different form 1.00 at the 5% level so that there is not evidence that the families are consuming more or less than their requirement in calories.

Statistical tests and estimation for the seven-day method (Model II)

We estimate v as $v^* = 1.18$, and var $(E_j)^* = 4.1350$, an estimate of the variance of each E_j. A further test shows that v^* is not significantly different from 1.00 at the 5% level. Although 1.18 would seem to indicate that the families are consuming 18% more calories than are required, the quantity is not significantly different from 1.00 at the 5% level, possibly because of the large variance of each E_j.

Table 4. Analysis of Covariance I

Source of variation	Degrees of freedom	Sum of squares for calories	Sum of squares for requirements	Sum of squares for cal. and req.	Adjusted degrees of freedom	Adjusted sum of squares	Mean adjusted sum of squares	F	F for 5% value
Day effects									
$d_1, d_2, \ldots, d_7$	6	472.81841	314.02932	354.39422	6	72.87496	12.14583	.68	2.40
Error	33	1991.36327	1107.25386	1254.12529	32	570.88482	17.84015		
Totals	39	2464.18168	1421.28318	1608.51951	38	643.75978			

The day effects are not significant at the 5% level.

Table 5. Analysis of Covariance II

Source of variation	Degrees of freedom	Sum of squares for calories	Sum of squares for requirement	Sum of squares for cal. and req.	Adjusted degrees of freedom	Adjusted sum of squares	Mean adjusted sum of squares	F	F for 1% value
Family effects									
$b_1, b_2, \ldots, b_{10}$	9	1255.26373	735.48471	867.71813	9	232.71822	25.85758	4.23	2.78
Error	54	522.72888	168.39804	183.04472	53	323.76356	6.10875		
Totals	63	1777.99261	903.88275	1050.76285	62	556.48178			

The family effects are significant at the 1% level.

Table 6. Analysis of Covariance III

Source of variation	Degrees of freedom	Sum of squares for calories	Sum of squares for requirement	Sum of squares for cal. and req.	Adjusted degrees of freedom	Adjusted sum of squares	Mean adjusted sum of squares	F	F for 5% value
Day effects									
$d_1, d_2, \ldots, d_7$	6	43.40457	5.24979	9.02214	6	29.93027	4.98838	.82	2.29
Error	54	522.72888	168.39804	183.04472	53	363.76356	6.10875		
Totals	60	566.13345	173.64783	192.06686	59	353.69383			

The day effects are not significant at the 5% level.

The corrected seven-day method (Model III)

The importance of the effect of family is demonstrated in table 5 which shows an observed $F = 4.23$, significant at the 1% level. The importance of the day effect is demonstrated in table 6 which has an observed $F = .82$, not significant at the 5% level. It would thus appear that the family effect is quite important while that of day is small in importance.

The velocity is estimated by $v^* = 1.09$. As in the other examples, the people apparently are consuming more calories than are "required", although the test of $v = 1.00$ is not significant at the 5% level. The variance of each E_{js} is estimated by 6.10875.

Conclusions about Acatenango

1. Day of the week is of very little importance in caloric consideration. This supports the frequent assertion that in a poor community the diet does not change from day-to-day. Of course, this applies only to caloric intakes, which we considered here, where, for example, the family may buy some meat over the weekend after the week's pay is received, but because they have animal protein calories, they lower their tortilla (corn) calories, keeping total calories the same, yet altering their animal protein intake.

2. Family is of great importance in caloric considerations. Of course, the effect of family includes many aspects such as education level, economic rank, and cultural foundation. The people that actually performed the dietetic survey claim that it is probably the economic facility that causes the dietetic differences in families, but this work is still in process with the collaboration of the antropological team.

3. The one-day method should not be used in practice because it does not allow for an effect of family, which has been shown by Model III to be of great importance. Furthermore, the intakes must be recorded daily for a family studied over the period of a week, lest the effects of day and of family be lost. Since, as far as calories are concerned, day of the week is not important, then it would seem that one could eliminate studying a family all week, and just be satisfied with two or more days, so that family effect could be seen, at the cost of losing some day effect. If, however, some other intake is considered, like that of animal protein, the study probably would have to be over the entire week.

There are other considerations against the one-day method. For instance, dietitians of INCAP who have worked in the field are always in agreement that many families lie on the first days of study. When the study is over a period of several days, the assertion is made that toward the end, the family does not bother to lie and that many of the lies, deletions, and errors of the first days can be corrected. It is not possible to demonstrate statistically whether families lie or not because Models I and II do not show a family effect, and in Model III the corrections that were made in the field were not recorded. Further work would have to be done to statistically test the assertion that the families lie significantly on the first days of study.

4. There is an indication that the families consume calories at a higher rate than "required" although the observed rate is not significantly different from the required at the 5% level.

Summary

In Acatenango, Guatemala studies were made of the dietetic intakes and requirements of groups of families, in the following way: 40 families were studied for exactly one day and a different group of 10 families was studied for exactly one week. This paper considers what information can be gotten from the caloric data by removing the "effect of requirement", points out important differences in the two types of studies, and shows how the "one-week method" was improperly performed, but indicates how this data was able at least partially to be salvaged. In particular, if it is important to see an "effect" of family, then it is not satisfactory to perform the one-day method, and if the seven-day method is used, the intakes must be recorded day-by-day and not just averaged over a week, lest the effects of family and of day be lost.

It is demonstrated that the effect of family is extremely important in Acatenango and that the effect of day of the week on caloric consumption is of very little importance.

References

1. Boletín de la Oficina Sanitaria Panamericana, Suplemento No 1, 119–129 (1953). — 2. STEEL, R. G. D. and J. H. TORRIE, Principles and Procedures of Statistics (London-New York 1960).

Authors' address:

Dr. CHARLES DEWITT ROBERTS and Miss MARINA FLORES INCAP, Guatemala, C.A.
N. Y. U. Graduate School of B. A., 100 Trinity Place, New York, New York 10006

From the Department of Biochemistry and Nutrition, Polytechnic Institute, Copenhagen (Denmark)

The relationship between diet and composition of bladder bile in mice

By H. DAM, F. CHRISTENSEN and I. PRANGE

With 5 tables

(Received August 22, 1968)

A diet without added fat induces abundant formation of cholesterol gallstones in young hamsters when the carbohydrate is furnished entirely or mainly in the form of glucose, whereas the tendency to production of gallstones is nil or very slight when the carbohydrate is furnished entirely in the form of rice starch. The different influence of the two carbohydrates with respect to gallstone formation in hamsters is paralleled by a difference in the composition of the bladder bile of that species; the ratio between lipid-soluble