Cohort Analysis: Analyzing Data on Product Use

Dr. Joseph L. Bonnici, Marketing, Bryant College
Dr. William B. Fredenberger, M.I.S., Valdosta State College

Abstract

Palmore's cohort analysis technique is a practitioner-oriented method to unravel age, period, and cohort effects in the data on product use. The technique is fully explained, and towards the end of the article, it is applied to grocery data for illustration purposes.

Introduction

Dissatisfaction with research based exclusively on cross-sectional data dates back to the eighteenth century (Baltes 1968). As early as 1741, Sussmilch, one of the first demographers, complained of the limitations of cross-sectional data. Yet, in spite of these limitations, nineteenth century psychologists used cross-sectional research techniques to pioneer aging research (Quetelet 1835, Galton 1883). It was early in the twentieth century that scientists like Darwin, Preyer, and Scupin introduced the idea of observing individuals over a period of time rather than just once. This new approach was considered as a biographical tool, and over the years longitudinal research became valued as an important supplement to cross-sectional data. It is uncertain who gave cross-sectional research and longitudinal research strategies their current names. In 1910, Camerer referred to the two methods of analyzing age relationships by two distinct names, "generalisierende" and "individualisierende". In 1931, Anderson wrote that the technical terms, "longitudinal" and "cross-sectional", were recent. However, Anderson failed to support his statement with any references. Slowly, the two methods made their way into other areas such as sociology (Mannheim 1952).

The two methods, by now referred to by the joint name, "cohort analysis", took a long time to gain recognition in marketing literature (Reynolds and Rentz 1981). Once they did, they were regarded as a promising new method of market research (Prester 1989, Rentz, Reynolds and Stout 1983) which could offer "human meaning of social change" (Campbell and Converse 1972).

The word "cohort" is derived from the Latin word "cohors" originally referring to "a division of a Roman legion." In cohort analysis, cohort refers to a group of persons that share a common event-origin within a given time period. French demographers, for example, use the French words "génération" for a cohort of births and "promotion" for a cohort of marriages (Wunsch and Termote 1978).

Imported into Marketing

By the late 1980's, cohort analysis gained a significant foothold in three different marketing areas. First, marketers see in it a way of segmenting the market. For example, Plummer (1990) uses it to divide the marketplace into mature Americans, baby boomers, and the next generation. Burnett (1989) uses cohort analysis to analyze the efficacy of retirement as a distinct segmentation variable. Gunnerson (1986) discovers through this analysis that consumers aged over fifty have much more discretionary income than those under the age of thirty-five, and that those aged between fifty-five and sixty-four enjoy the highest per capita income in the country.

A second area for cohort analysis is advertising, where success is directly dependent on targeted segments. For example, Norvell (1988) describes the advertising approach towards older consumers whose perception of a product's potential evolves over time. Exter (1986) suggests a "multidimensional view of age" in designing advertising campaigns. This is supported by Day, Davis, Dove and French (1987/88) who offer advertising guidelines for four senior citizen market segments.

The third significant area is demand modeling and forecasting. For example, lifestage marketing examines the booms and busts of various age groups, as well as their needs and wants during particular life stages (Ambery 1990). A Swedish model of car ownership reveals that age rather than income is the leading factor in forecasting ownership (Jansson 1989). Another study shows that older people are active shoppers and that over time they are more and more using electronic media for market information purposes (Tongren 1988).
Basic Cohort Table

For cohort analysis to occur it is necessary to measure one or more cohorts at two or more separate times (for example, with a ten-year gap between them). One may want to question whether consumers in Cohort Y changed their pizza consumption after ten years. Or one may want to know whether Cohort Z buys less computers than the younger Cohort X. The behavior examined could be very extensive and may include consumption, reading, voting and working (Wood 1986).

Table 1 illustrates a basic cohort table based on a hypothetical example of percentage of readers subscribing to a magazine. A bird's eye view of the numbers in the table is unlikely to reveal anything other than that we are measuring four age groups at four points in time (1975, 1980, 1985, 1990). However, once the eye trains itself in three particular directions, patterns of subscription start to emerge.

Table 1
A Cohort Table of Percentage of Readers Subscribing to A Hypothetical Magazine
(Ten-Year Cohorts Measured at Five-Year Intervals)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>60-69</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>50-59</td>
<td>0.08</td>
<td>0.08</td>
<td>0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>40-49</td>
<td>0.10</td>
<td>0.20</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>30-39</td>
<td>0.25</td>
<td>0.21</td>
<td>0.18</td>
<td>0.10</td>
</tr>
</tbody>
</table>


If one looks horizontally, and compares the age groups, the 30-39 year age group offers the most promising segment for the advertisers in the magazine. It has consistently outperformed the other three age groups at the subscription department. Hence, by examining the rows it is possible to detect age effects. It is pertinent to point out that the age effects in the table refer to chronological aging. Only chronological aging is taken directly into account in cohort analysis (Glenn 1977).

A vertical reading of the table leads to contrasts and comparisons between the time periods. This results in period effects. It seems that 1990 is the worst period in the table with subscriptions hitting an all-time low in three of the age groups. On the average, based on the limited amount of data available, subscriptions seem to have reached a plateau around the 1980 and 1985 data, only to slip away in 1990. There is one exceptionally bright spot in 1990. Never before have so many aged 60-69 years subscribed to the magazine. Why does this jump occur?

A diagonal reading of the data possibly explains why. The diagonal reading reveals seven cohorts of distinct birth cohort membership represented by the edges of the table. C6, seen as a cohort, is one of the best cohorts in the magazine's history as suggested by the table. This cohort has been relatively avid in subscribing. It seems to include a core of loyal readers. In 1990, this cohort is represented by the 60-69 year age group. This cohort offers the most plausible explanation for the aforementioned exceptionally bright spot in 1990 - the 7 percent of the 60-69 year group. Cohort effects, revealed by tracing the cohort groups diagonally show that readership declines over time for each cohort group. Beyond C6, the future ratings of the younger cohorts (C5, C4 and C3) will probably continue to shrink. The subscription behavior of these latter generations suggests less than a promising horizon for the magazine.

Such an analysis is possible because Table 1 has been constructed in a way where the magnitude and direction of the hypothetical percentage numbers selected facilitate a visual interpretation. Glenn (1977) supports such an "eyeball" interpretation by noting that a purely mechanical cohort analysis is a waste of time and should be avoided. He suggests that a cohort analyst should be able to visually examine a cohort table. Although there is truth in Glenn's opinion, visual interpretation is rarely possible in real life since cohort tables always contain a minimum level of confounding. Blalock (1966; 1967) calls this confounding "the identification problem." The problem with cohort tables is that regardless of the direction one reads the data, two of the three effects (period, age, and cohort) are confounded. This is due to the fact that the period of time is linearly dependent on cohort membership and age, the cohort membership is linearly dependent on period and age, and the age is linearly dependent on period and cohort. This means that if we know how old a subscriber was in a particular year, then we can perfectly predict in which cohort he is a member. In other words, information about the subscriber's locus on two of the elements automatically guarantees information on the third element. Hence, it does not make sense to use a regression analysis to
analyze cohort data since the three elements cannot be simultaneously used as independent variables (Glenn 1977).

How can one then separate age, period, and cohort? Several methods have been proposed, other than the visual inspection method. Techniques have been suggested to unravel the effects. These techniques include Iversen and Norpoth's (1976) analysis of variance, Goodman's (1972) log-linear model, Aigner's (1973) and Klecka's (1971) estimates of period effects, Mason, Mason, Winsborough, and Poole's (1973) multiple classification analysis, Schaefer's (1965) sequential strategies model, and Baltes' (1968) bifactorial model.

These research methods are based on assumptions about the existence, consistence and directionality of cohort, period or age effects. These assumptions are necessary because of the confounding elements in the observable differences. Yet, at most, each methodology is as valid as the assumptions on which it is based. Most cohort analyses seek to detect the presence of cohort, period or age effects and to find out their magnitude and directionality. Rarely does it make sense to assume beforehand the value of any of the three effects. This is where the above techniques falter. For example, Baltes' bifactorial model assumes that period effects are zero, and Schaefer assumes that period effects are positive. In fact, if a cohort analysis methodology is used under such rigorous assumptions, the methodology's appropriateness is limited for most problems (Rentz 1980).

**Triad Method**

It is for these reasons that we propose another cohort analysis methodology. It is Palmore's (1978) triad method. Rather than cutting right through the data and trying to come up with a biased answer out of the midst of the confounding, Palmore's method "peels off" the data into three layers or levels, and makes sense out of them in a dependent fashion. The method is logically appealing, simple, and offers a methodologically adequate tool for the marketing practitioner. It requires no prior knowledge of statistical methods other than a basic understanding of t-test statistical techniques.

The method's first layer consists of measuring three differences (Figure 1):

1. **Longitudinal** is the difference between two measurements of the same cohort over time. In mathematical terms, this is the subtraction of Cell X from Cell Y.
2. **Cross-sectional** difference occurs between younger and older cohorts at a specific time. It is the difference between Cell Z and Cell X.
3. **Time-lag** difference occurs between two cohorts measured at two different points in time such that their age is equal at the respective times of measurement. This age difference constrains the timing of the measurement. Mathematically speaking, the time-lag difference is the difference between Cell Y and Cell Z.

![Figure 1](image.png)

Adapted from E. Palmore (1978). "When Can Age, Period and Cohort be Separated?" Social Forces, 57 (1), 284.

After the three differences are calculated, one can proceed to the second level of analysis by exploring the two basic constituents of each difference. Simply stated, each difference is made up of two effects, as follows:

- Longitudinal difference = Period + Age
- Cross-sectional difference = Cohort + Age
- Time-lag difference = Period - Cohort

Thus, if a group of retired people consumes more aspirin today than they did twenty years ago, this longitudinal difference may be due either to the evolving time or period in which they live (for example, based on recent discoveries, physicians are encouraging an increase in aspirin consumption as a way to reduce the incidence of heart attacks) or to their age (the older people get, the more they are likely to require medication) or to both period and age. Thus every significant difference automatically leads to a scrutiny of its pair of effects.

If no significant differences occur, then there are no period, age or cohort effects. In other words if consumption across cohorts remains the same for all age
groups across time, the three differences (longitudinal, cross-sectional, and time-lag) are zero. Example 1 represents a pattern of no significant differences.

Example 1

**NO EFFECTS UNDER SIGNIFICANT DIFFERENCES**

<table>
<thead>
<tr>
<th>Per Capita Monthly Consumption of Vitamin XYZ Capsules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Year</td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td>1940</td>
</tr>
<tr>
<td>1930</td>
</tr>
<tr>
<td>Longitudinal difference</td>
</tr>
<tr>
<td>Cross-sectional difference</td>
</tr>
<tr>
<td>Time-lag difference</td>
</tr>
</tbody>
</table>

If only one significant difference occurs, there probably is an error in the data. A look at the above difference equations shows that one significant difference is theoretically impossible. If such a situation occurs, there is an error in the data.

When two significant differences occur, chances are that there is one pure effect behind them. To find out which effect it is, one should check which effect is common to the significant pair of differences:

Cross-sectional and Longitudinal = Age effect  
Cross-sectional and Time-lag = Cohort effect  
Time-lag and Longitudinal = Period effect.

Example 2 illustrates two significant differences where the period effect is common to both of them. However, there may be a much more complex possibility under two significant differences, when there are two equal and opposite effects. This requires a complex interpretation. Palmore suggests that unless there is other evidence to the contrary, the principle of parsimony justifies the interpretation of two significant differences as emanating from one effect.

When three significant differences occur, it is impossible to estimate the value of each of the three effects unless there is reason to assume that one of the effects is equal to zero. If such an assumption is possible, then one can pin down the value of the other two effects. Example 3 illustrates such a case where the longitudinal, cross-sectional and time-lag differences are significant. However, suppose that on closer analyses one finds that there was a similar amount of rainfall across the state of Alabama during the two years under study. If this leads the researcher to assume that period effects were not responsible for any changes in sprinkling, then the researcher can proceed and deduce the magnitude of age and cohort effects.

Example 2

**PERIOD EFFECT UNDER TWO SIGNIFICANT DIFFERENCES**

<table>
<thead>
<tr>
<th>Per Capita Monthly Usage of XYZ Credit Card</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Year</td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td>1944</td>
</tr>
<tr>
<td>1939</td>
</tr>
<tr>
<td>Longitudinal difference*</td>
</tr>
<tr>
<td>Cross-sectional difference</td>
</tr>
<tr>
<td>Time-lag difference*</td>
</tr>
</tbody>
</table>

* t-test statistically significant

The separation of cohort, period and age effects necessitates the third and final level of analysis. In Palmore’s (1978, p.286) words:

"The problem still remains of imputing causes for these effects. Age effects may be produced by any combination of biological aging, atrophy caused by inactivity, aging of cognitive processes, movement to different age-related roles, age discrimination, etc. Period effects may be caused by changing physical...environments, changes in measurement techniques or group composition, practice effects due to exposure to the measure, and the like. Cohort effects may be caused by historical differences in social or physical environments during critical earlier years, genetic differences between cohorts [and] differences in size or structure of cohorts."

To deduce which causes are leading to the effects, one has to look for experimental, historical and other outside evidence. Hence, it is important to research the possible causes thoroughly and understand the directionality of their impact.
Example 3
TWO EFFECTS UNDER THREE SIGNIFICANT DIFFERENCES

Per Capita Monthly Sprinkling of Lawns in Alabama

<table>
<thead>
<tr>
<th>Birth Year</th>
<th>Year 1971</th>
<th>Year 1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>1941</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1921</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Longitudinal difference* = 3 = Period (0) + Age (3)
Cross-sectional difference* = 2 = Cohort (-1) + Age (3)
Time-lag difference* = 1 = Period (0) - Cohort (-1)

*t-tests statistically significant

Application: Grocery Data

To illustrate the ease and feasibility of Palmore’s cohort analysis method, we apply the method to data about grocery products’ usage. The data were collected by Simmons Market Research Bureau (SMRB) in 1979 and in 1989. The respondents were over 15,000 adults in each study. Out of the whole data base, we examined the usage at home of 68 grocery products. We focused on two cohorts. The first cohort is represented by adults who were 55-64 years in 1979. The second cohort is represented at two points in time: by adults who were 45-54 years in 1979, and in 1989 when the same cohort’s age was 55-64 years. Table 2 represents each cohort’s average percentage of users of the grocery items at the specific points in time. Using a t-test (α =.05) on the difference between percentages, it is found that the percentage registered for the 55-64 age group in 1979 (i.e. 36.10882) is not significantly different from that registered for the other cohort group aged 55-64 ten years later (i.e. 34.65441). Hence, the time-lag difference is considered as not significantly different from zero. However, according to the t-test (α =.05), both the longitudinal difference and the cross-sectional difference are significantly different from zero.

This leads us to the second level of analysis where:

Longitudinal difference = Period + Age
Cross-sectional difference = Cohort + Age

Age is the common element. Hence, the pattern of two significant differences is due to the age effect. Since the period effect and the cohort effect do not lead to a significant difference in the time-lag difference, we do not expect them to result in a significant difference when paired with age. This means that the proportion of the 55-64 age class who used the grocery items in 1989 was more likely to be similar to the proportion of the 55-64 age class of 1979. Thus, there is an age effect on the usage of grocery products. In this case, one cannot determine with absolute certainty that the age effect is a pure or exclusive effect, because the computed longitudinal difference is not exactly equal to the cross-sectional difference. Such difference between the longitudinal and the cross-sectional differences is typical of accuracy problems in self-report measurements (Henerson, Lyons and Taylor 1987).

Palmore’s final level of analysis necessitates a justification for the age effect. This is not difficult in our case because scientific studies show that people need less caloric intake as they get older. Hence, it is no surprise that we witness a common experience between the 55-64 age classes of 1979 and 1989 in their grocery products usage, and a significant consumption difference between these age classes and the 45-54 age class. As expected, the 45-54 age class scores significantly higher on consumption.

Usage Recommendations

There are various reasons why one should subject such type of data to cohort analysis. First, cohort analysis works extremely well with secondary data. This saves both money and time. Second, Palmore’s method is so simple to use that a marketer can use it without having to recruit market research staff. Third, the widespread access to data archives around the United States eliminates a lot of data collection problems. A fourth advantage is that subjecting data to cohort analysis can alert the marketer to underlying trends, even at an early stage (Kiecolt and Nathan 1985).

The application of the method to marketing problems also makes sense because consumption may be dependent on people and their surroundings over a time period. Such rationale lead Time Inc. to apply cohort analysis to its readership data from 1965 to 1985. What it discovered in the process was that period effect was most pronounced during Watergate, around 1974. It also found that Baby Boomers as a cohort group "didn't stick it to Time" (Wood 1986).

Summary

Conventional research based only on cross-sectional data is inadequate. It offers an incomplete analysis.
Table 2. PERCENTAGE OF COHORTS USING THE GROCERY PRODUCTS

<table>
<thead>
<tr>
<th>Ages</th>
<th>1979</th>
<th>Year</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>45-54</td>
<td>39.88824</td>
<td></td>
<td>34.65441</td>
</tr>
<tr>
<td>55-64</td>
<td>36.10882</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>34.65441 - 39.88824 = -5.23383*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. error of difference</td>
<td>0.96354</td>
</tr>
<tr>
<td>95% C.I. of difference</td>
<td>-7.15704</td>
</tr>
<tr>
<td>T value</td>
<td>5.43188</td>
</tr>
<tr>
<td>Probability level</td>
<td>0.0000</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.9542</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CROSS-SECTIONAL DIFF.</th>
<th>36.10882 - 39.88824 = -3.77942*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. error of difference</td>
<td>0.54543</td>
</tr>
<tr>
<td>95% C.I. of difference</td>
<td>-4.86809</td>
</tr>
<tr>
<td>T value</td>
<td>6.92921</td>
</tr>
<tr>
<td>Probability level</td>
<td>0.0000</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.9849</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TIME-LAG DIFFERENCE</th>
<th>34.65441 - 36.10882 = -1.45441</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. error of difference</td>
<td>1.05289</td>
</tr>
<tr>
<td>95% C.I. of difference</td>
<td>-3.55598</td>
</tr>
<tr>
<td>T value</td>
<td>1.38135</td>
</tr>
<tr>
<td>Probability level</td>
<td>0.1718</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.9370</td>
</tr>
</tbody>
</table>

* Significantly different from zero at α = .05 using a t-test on the difference between proportions.

The addition of longitudinal data enriches the research methodology by providing a broader picture of consequences and possible causes.

Cohort analysis as a comprehensive technique suggests three possible causes or effects, acting singly or in combination. Age effects refer to biological and psychological changes experienced in chronological aging. Period effects have to do with environmental changes. Cohort effects refer to genetic change across history.

Although there are several statistical techniques seeking to unravel the three possible effects, most techniques are hampered by rigorous assumptions. Palmore's method offers a unique alternative on the basis of clarity of thought and method. Proceeding in a "peeling off" fashion, it suggests three levels of analysis which between them direct the cohort analyst's attention to the "what? where? when? why?" aspects of the issue at hand. The method is easy to understand. It is also highly practical as demonstrated in the above application.

Suggestions for Future Research

Palmore's cohort analysis technique and its application to the grocery data demonstrate the relevance of cohort analysis to consumer behavior. Cohort analysis suggests that we step back from aggregate consumption and analyze such consumption in terms of subgroups based on the ageing process, cohort membership, and particular events surrounding the consumption experience.

Rentz (1980) suggests that cohort analysis is a disaggregative tool and recognizes several demand sched-
ules where only one was recognized before. Therefore, cohort analysis may also be beneficial when applied to strategic planning processes, in anticipation of change. It is also relevant to public policymakers seeking a better understanding of gerontological developments.

The authors are indebted to the anonymous JABR reviewer for the useful and considered suggestions, and to Mrs. Barbara Cape for the helpful assistance on this paper.

***References***

1989.