1989 Cultural Anthropology Methods Newsletter 1(1):3-5

# Problems in Determing the Audience Composition of a Multi-Day Festival from Survey Data

*John Gatewood (Lehigh), Catherine Cameron (Crest College) & Bennet Eisenberg (Lehigh)*

—

Each August since 1984, a nine-day music festival has been held in Bethlehem, Pennsylvania. The extravaganza, called "Musikfest," attracts crowds totaling nearly half a million people and is run by an independent nonprofit organization. Local banks and businesses sponsor the festival through contributions of money and/or personnel, and more than a thousand private citizens volunteer their time and energies to make Musikfest a success (see Cameron, 1987, for a fuller discussion).

From the beginning, festival organizers saw the benefits of conducting visitor surveys. Audience composition — in terms of simple demographic variables such as residence, age, sex, and income — is used by the marketing committee to attract corporate sponsors. We have conducted these surveys over the past four years and have encountered problems with what appeared to be a straightforward matter. These problems and our solutions are what we would like to share with readers.

The question we were trying to answer is deceptively simple: "What is the proportion of local versus nonlocal individuals who come to Musikfest?" (This is an important question because we use this information in conjunction with per capita expenditure rates and total crowd size estimates to assess the festival's economic impact.) The first item in our questionnaire asks respondents where they live, so it would appear a simple tally of responses would provide the answer. Unfortunately, such a tabulation misses the mark, for it does not take into account a couple of sampling issues.

If the groups' attendance patterns were the same, then the percentages observed in the sample would accurately reflect, within the normal bounds of statistical inference, the proportions of local and nonlocal individuals who attended the festival. But, when groups differ in attendance patterns, the percentages obtained in a sample must be adjusted according to the differential likelihood of locals and nonlocals being interviewed. Before illustrating this with real data, let us first consider some hypothetical situations.

---

EXAMPLE 1. Suppose there were a population of 1000 Smiths and another of 1000 Browns. A ten-day festival is held, and each person signs a guest register when he or she first arrives, i.e., once. Each Smith attends all ten days of the festival (producing 10,000 Smith-attendances), but each Brown attends only five days (producing 5000 Brown-attendances). After the first day of the festival, all 1000 Smiths will have signed the guest register. The Brown sign-in will take longer, but by the sixth day all 1000 Browns will have signed; hence, the final guest register will show 50% Smiths and 50% Browns.

If each Brown's pattern of attendance is independent of his or her fellows, then roughly 500 Browns will be at the festival on any given day, whereas all 1000 Smiths will be there. Thus, the sum of random daily samples would show roughly 67% Smiths and 33% Browns in the festival's audience, yet the guest register would show that the same number of Browns attended as Smiths.

EXAMPLE 2. Again, suppose there were two populations of 1000 individuals apiece, another ten-day festival, and another guest register. This time, each Smith and Brown attends only five days of the ten-day event. By the sixth day, all 1000 Smiths and all 1000 Browns will have signed the guest register, resulting in 50%-50% proportions.

Now, it could happen that all the Smiths attend the first five days of the festival and all the Browns attend the second five days. It could happen that all the Smiths attend on odd-number days, and all the Browns attend on even-number days. It could happen that 500 of the Smiths attend each day, and likewise the Browns. So long as the two groups' patterns of attendance are similar, the daily crowd size may fluctuate, but the sum of daily random samples will have roughly 50% Smiths and 50% Browns, the same as shown in the guest register.

If, however, the attendance patterns of the two groups are dissimilar, then the sum of daily samples will deviate from the proportions shown in the guest register. For example, if the Smiths appear 500 strong each day, but the Browns, all 1000 of them, attend only on odd-number days, then the sum of daily random samples of fixed size1 would show roughly 67% Smiths and 33% Browns instead of the 50%-50% reflected in the guest register (see Table 1).

---

The two general kinds of sampling effects we should like to extract from the above examples are as follows.

SAMPLING EFFECT #1: "Groups attend different number of days." The proportions of Smiths and Browns (or locals and nonlocals) in a sample reflect the proportions of attendances produced by each group, not the proportions of attendees. As Example 1 shows, sample data will conform to "guest register" proportions only when the number of attendances per person is the same for the groups. Thus, to estimate the proportion of attendees (Smiths vs. Browns, locals vs. nonlocals), we must adjust the sample percentages using

Table 1. The effect of even versus sporadic attendence on sample percentages

| | Days of Festival | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Sum | % |
| *Daily Attendances* | | | | | | | | | | | | |
| Smiths | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 5,000 | 50% |
| Browns | 1,000 | 0 | 1,000 | 0 | 1,000 | 0 | 1,000 | 0 | 1,000 | 0 | 5,000 | 50 |
| *Sample Data (12 respondents each day of festival)* | | | | | | | | | | | | |
| Smiths | 4 | 12 | 4 | 12 | 4 | 12 | 4 | 12 | 4 | 12 | 80 | 67 |
| Browns | 8 | 0 | 8 | 0 | 8 | 0 | 8 | 0 | 8 | 0 | 40 | 33 |

each group's mean number of attendances per attendee (i.e., days at the festival).

SAMPLING EFFECT #2: "Groups' daily attendance patterns differ." As Example 2 shows, even when groups produce the same number of attendances per person, how these attendances are distributed over the duration of the festival (the daily variation) must be taken into account. So long as the groups exhibit similar patterns, this second sampling effect can be ignored. In general, however, the group whose attendance pattern has the greater variance (is more sporadic) will be underrepresented in the final sample.

In our work with Musikfest, we encounter very directly the first sampling effect. Last year, for instance, our survey found that 323 respondents (74.1%) resided locally, and 113 respondents (25.9%) were from out of town. Locals spent an average of 4.577 days at the nine-day festival, whereas nonlocals attended an average of only 2.505 days. Since there was no significant difference between the groups in terms of daily attendance patterns, we could ignore the second sampling effect, but we must adjust the sample percentages of locals and nonlocals because the two groups differ very significantly in terms of the number of attendances per attendee.

The observed percentage of local respondents in the sample should, within normal statistical expectations, be equal to the number of local-attendances divided by the total number of attendances (local-attendances plus nonlocal-attendances), as expressed in equation (1). The case for nonlocals is similar, as expressed in equation (2).

Locals: $$\frac{ax}{ax + by} = p \qquad (1)$$

Nonlocals: $$\frac{by}{ax + by} = q \qquad (2)$$

where

x = actual (but unknown) number of local attendees

y = actual (but unknown) number of nonlocal attendees

a = mean number of attendances per local attendee

b = mean number of attendances per nonlocal attendee

p = observed percentage of local respondents in the sample

q = observed percentage of nonlocal respondents in the sample

From equations (1) and (2), we may derive the simplified formulas for calculating adjusted proportions of local and nonlocal attendees, that is, $x/(x+y)$ and $y/(x+y)$, respectively.

Locals: $$\frac{x}{x + y} = \frac{bp}{a - ap + bp} \qquad (3)$$

Nonlocals: $$\frac{y}{x + y} = \frac{aq}{b - bq + aq} \qquad (4)$$

Plugging last year's Musikfest data into equations (3) and (4), we find that the sample percentages—74.1% locals and 25.9% nonlocals—adjust to 61.0% and 39.0%, respectively. In other words, although nonlocal residents made up only about one-quarter of the crowd size any given day of the festival, a guest register would have shown that about two out of five of the individuals who attended Musikfest '88 were out of towners. Table 2 shows the magnitude of these adjustments for each of the past four years of Musikfest.

To illustrate concretely the difference this adjustment can make, we turn now to a consideration of the induced spending associated with Musikfest. While spending of any sort helps the area's service industries, the local economy benefits most from tourist dollars. Thus, the amounts and proportions of local versus nonlocal spending is a matter of great concern as festival organizers market the event to corporate sponsors and negotiate with the city government over parking, street usage, and extra police costs.

The "normal" way of estimating induced spending from audience survey data uses whole-sample means and the unadjusted sample percentages, as follows (Method 1).

1. Find the audience size: divide the total crowd size by the whole-sample's average number of days people attended the festival.
2. Partition the audience into local and nonlocal residents: multiply the total audience size times the sample's percentages of locals and nonlocals, respectively.
3. Find how much money was spent by each group: multiply the whole-sample's average per capita spending times the number of locals and nonlocals, respectively.
4. Compute total audience spending: add the amounts spent by locals and nonlocals.

Given our previous arguments, however, we know that Method 1 will underestimate the amount of tourist dollars generated by Musikfest. To adjust for sampling effect #1, we should estimate induced spending as follows (Method 2).

1. Partition the crowd size into attendances produced by locals and attendances produced by nonlocals: multiply the total crowd size times the unadjusted sample percentages of locals and nonlocals, respectively.

2. Find how many local and nonlocal residents were in the audience: divide the number of local-attendances by the average number of days locals attended, and divide the number of nonlocal-attendances by the average number of days nonlocals attended.

3. Find how much money was spent by each group: multiply the number of locals times their average per capita spending, and multiply the number of nonlocals times their average per capita spending.

4. Compute total audience spending: add the amounts spent by locals and nonlocals.

Method 1 (unadjusted) and Method 2 (adjusted) lead to substantially different conclusions about the economic impact of Musikfest, especially when per capita spending for locals and nonlocals differs, as happened last year. Figure 1 graphically illustrates the magnitude of these differ-

**Table 2. Adjustments to Musikfest data: 1985–1988**

| | 1985 | 1986 | 1987 | 1988 |
|---|---|---|---|---|
| **Locals** | | | | |
| **Sample %** | 86.3 | 78.5 | 74.1 | 74.1 |
| **Days/Local** | 4.656 | 4.798 | 4.811 | 4.577 |
| **Adjusted %** | 79.4 | 67.4 | 60.6 | 61.0 |
| **Nonlocals:** | | | | |
| **Sample %** | 13.7 | 21.5 | 25.9 | 25.9 |
| **Days/Nonlocal** | 2.856 | 2.718 | 2.583 | 2.505 |
| **Adjusted %** | 20.6 | 32.6 | 39.4 | 39.0 |

Legends mislabeled. Solid bars are Method 2.

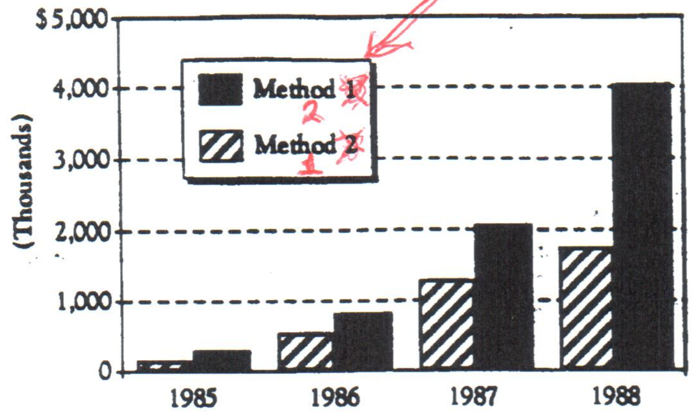


**Figure 1. Tourist spending: two methods contrasted.**

ences in terms of the key political issue of "tourist" dollars generated in the local economy as a result of Musikfest.

Although sampling effect #2 has not yet been a concern in our Musikfest surveys, we should like to explain how to adjust for it as well.

Say the numbers of Smith-attendances and Brown-attendances for the ith day of an event that last t days are $n_i$ and $m_i$, respectively. The total observed crowd size for the ith day is, thus, equal to $(n_i + m_i)$. Because we cannot know the daily crowd sizes in advance, however, our research is designed to sample a fixed number of individuals each day. Each daily sample, of constant size k, will be composed of some Smiths and some Browns, denoted $p_i$ and $q_i$, respectively, and $p_i + q_i = k$.

Presuming the samples are representative, then sample proportions for a given day will reflect the proportions of Smiths and Browns in that day's audience, as expressed in equations (5) and (6).

$$\text{Smiths:} \quad \frac{p_i}{k} = \frac{n_i}{n_i + m_i} \qquad (5)$$

$$\text{Browns:} \quad \frac{q_i}{k} = \frac{m_i}{n_i + m_i} \qquad (6)$$

Over the duration of the festival (i.e., for all t days), the true proportions of Smiths is $n/(n+m)$, where $n = n_1 + n_2 + \ldots + n_t$ and $m = m_1 + m_2 + \ldots + m_t$, and similarly for the proportion of Browns. As we noted in Example 3, however, the proportions computed from the sum of daily samples do not, in general, correspond to the true proportions, i.e., the left- and right-hand

$$\text{Smiths:} \quad \frac{p_1 + p_2 + \cdots + p_t}{tk} \neq \frac{n}{n+m} \qquad (7)$$

$$\text{Browns:} \quad \frac{q_1 + q_2 + \cdots + q_t}{tk} \neq \frac{m}{n+m} \qquad (8)$$

sides of equations (7) and (8) are usually unequal

Therefore, to make the sample data equal the true proportions, we must weight each day's sample results by the total audience size for that day, as shown in equations (9) and (10).

$$\text{Smiths:} \quad \frac{\frac{p_1}{k}(n_1 + m_1) + \frac{p_2}{k}(n_2 + m_2) + \cdots + \frac{p_t}{k}(n_t + m_t)}{n+m} = \frac{n_1 + n_2 + \cdots + n_t}{n+m} = \frac{n}{n+m} \qquad (9)$$

$$\text{Browns:} \quad \frac{\frac{q_1}{k}(n_1 + m_1) + \frac{q_2}{k}(n_2 + m_2) + \cdots + \frac{q_t}{k}(n_t + m_t)}{n+m} = \frac{m_1 + m_2 + \cdots + m_t}{n+m} = \frac{m}{n+m} \qquad (10)$$

If we do not specify weighting coefficients (i.e., if we treat the daily samples equally), this has the effect of inadvertently "over-weighting" those samples taken on small attendance days and "under-weighting" samples taken on large attendance days. Of course, when Smiths and Browns have similar daily attendance patterns, then no weighting is necessary, because in that special case $n_i/(n_i + m_i)$ is the same for each day and equal to $n/(n+m)$.

In conclusion, we hope to have: (a) drawn attention to two sampling problems in survey research, which we suspect are generally overlooked, (b) explained ways of dealing with them, and (c) illustrated the differences they make in a real-life situation of some importance. We welcome reader comments on these matters.

*References*


Cameron, Catherine M. 1987. The Marketing of Tradition: The Value of Culture in American Life. *City and Society*, 1(2):162-174.


# Hypertext

*(continued from page 12)*

families, in the colonias of Tijuana, as part of their undergraduate field training course. Notes are broken into short, meaningful chunks with header information (name, date, time, place, persons present, topics, etc.), and sequential jumps to keep them in chronological order. From the header information we cross-reference the files into a coherent data base. As students and instructors use the files, they can edit the conceptually organized files that we use to find comparable topical and other information across the fieldnote files. The original files are thus relatively static, except for editing to improve the information in the headings, but the other files — our current conceptualization of the data base — are in constant evolution. Students keep their own fieldnote files on diskettes, but our cross-referenced copy of their notes remains with our cumulative archive. Thus, if someone using hypertext wants to "go down a particular street in the colonia" to see what each student field worker has reported about the tasks that children do in their household, hypertext both finds the relevant households (pulling up a map of the colonia in graphic mode) and the relevant topical information.

Sample hypertext data bases are available from Neil Larson or, for anthropological applications, through the *World Culture Journal*. Larson seems always happy to talk to those interested ([415]428-0104).