Schadenfreude is launching a new insurance product: one-time policies for concert promoters, each covering a single concert and paying a fixed benefit in the event of the concert’s cancellation. In this report, we describe the calculation of fair value and premium cost for such a policy. For purposes of illustration, we will assume that the policy covers a polka concert scheduled in Alaska, with a $30,000 benefit paid in the event of cancellation; but the mathematics are valid for every combination of genre, location, and benefit value.

1 The policy

The policy we are offering pays a fixed benefit (in our example, $30,000) in the event that the concert is cancelled due to forces beyond the promoter’s control. Such circumstances include, but are not limited to, earthquake, fire, tsunami, or flood; prolonged power failure; denial of access to the venue for safety or security reasons; or non-appearance of key performers. Cancellation due to weather is only covered if the policy is purchased at least 30 days before the date of the event.

The benefit will not be paid if the event is cancelled for reasons deemed to have been in the promoter’s control. These include, but are not limited to, insufficient ticket sales or sponsorship income; failure to make adequate advance arrangements for the appearance of the performers or the use of the venue; or financial failure of the promoter.

In dealing with the promoter, it should be emphasized that this policy provides no liability coverage for the event.

2 Hazards

Reasons for cancellation of an event can be broadly grouped into two categories: those related to the location, and those related to the performer.

Location affects the probability of cancellation. Different places are differently subject to the risks of weather, earthquake, power failure, epidemic disease, riot, insurrection, and terrorism.

The risk of cancellation also depends on the performer. Individual performers differ in their tendency to high-risk behaviors, in their susceptibility to disease and disability, and in their probability of being in jail when they should be onstage.

Unfortunately, the record of cancellation for a specific venue or for a particular performer is generally too small to allow an accurate assessment of the probability of cancellation of future events. To obtain a sample large enough to produce an accurate estimate of cancellation probability, we must use larger groupings.

Experience suggests that the risk of cancellation can be predicted with reasonable accuracy based on the region of the United States in which the performance is scheduled, and on the genre of the performance. We have access to two databases containing information on these. The first of these was compiled between 2002 and 2004 by the
Federal Minor Inconvenience Management Agency; it contains records of 13,864 events, indicating the region of the country in which each event was scheduled and whether or not it was cancelled. For brevity, we will refer to it in this report and the accompanying Excel file as “Data1”. The second database, hereinafter referred to as “Data2”, was compiled between 2002 and 2005 by the American Live Entertainment Institute. It lists 16,706 events, for each one indicating the genre of the performance and whether or not the event was cancelled.

Unfortunately, we do not have data records that simultaneously list region and genre. “Data1”, which breaks down concerts by region, says nothing about genre; “Data2”, while it contains information about genre, is mute on the subject of location. In order to estimate the cancellation probability for a specific combination of region and genre, we will need to make certain assumptions.

The two databases were compiled over roughly the same time period, and each is based on a nationwide survey of concert events. It is reasonable to assume that they describe similar sets of concerts, and that conclusions drawn from one database can be applied to the other.

3 Cancellation probability

Let $R$ be the event that a concert is scheduled for a particular region of the United States (in our example, Alaska). Let $G$ be the event that the performance belongs to a particular genre (in our example, polka). Let $C$ be the event that the concert is cancelled. To calculate the fair value of a cancellation policy, we need the probability that a concert of that particular genre in that particular region is cancelled:

$$P(C \mid G \cap R)$$  (1)

The standard formula for conditional probability allows us to write

$$P(C \mid G \cap R) = \frac{P(C \cap G \cap R)}{P(G \cap R)}$$  (2)

We will assume that $G$ and $R$ are independent events. Intuitively, this assumption may seem unwise. We are accustomed to thinking of musical tastes as varying with geography: people in Texas listen to country, people in Tennessee like bluegrass, people in Wisconsin prefer polka.

In fact, musical tastes are more geographically homogeneous than we might think. The influence of national radio and television networks, and the tendency of Americans to move cross-country, have greatly reduced these regional differences. There is still some geographic variation, but it is strongest among the oldest people—which is to say, those least likely to attend concerts. Our assumption of independence is not an unreasonable one.
Given this assumption, (2) can be rewritten as

\[
P(C|G \cap R) = \frac{P(C \cap G \cap R)}{P(G)P(R)}
\]  

(3)

We can rewrite the numerator of (3):

\[
P(C|G \cap R) = \frac{P(G \cap R|C)P(C)}{P(G)P(R)}
\]  

(4)

At this point, we will further assume that \(R\) and \(G\) are not only independent, but remain independent when conditioned on \(C\). This allows us to write our final expression:

\[
P(C|G \cap R) = \frac{P(G|C)P(R|C)P(C)}{P(G)P(R)}
\]  

(5)

All five factors in (5) can be estimated from “Data1” and “Data2”. For our example of a polka concert scheduled in Alaska:

\[
P(G|C) = \frac{\text{# cancelled polka concerts in “Data2”}}{\text{# all cancelled concerts in “Data2”}} = \frac{94}{1042} = 0.0902
\]

\[
P(R|C) = \frac{\text{# cancelled Alaska concerts in “Data1”}}{\text{# all cancelled concerts in “Data1”}} = \frac{33}{892} = 0.0370
\]

\[
P(C) = \frac{\text{# all cancelled concerts in “Data1” and “Data2”}}{\text{# all scheduled concerts in “Data1” and “Data2”}} = \frac{1934}{30570} = 0.0633
\]

\[
P(G) = \frac{\text{# scheduled polka concerts in “Data2”}}{\text{# all scheduled concerts in “Data2”}} = \frac{1064}{16706} = 0.0637
\]

\[
P(R) = \frac{\text{# scheduled Alaska concerts in “Data1”}}{\text{# all scheduled concerts in “Data1”}} = \frac{1026}{13864} = 0.0740
\]

Substituting these values in (5) yields

\[
P(C|G \cap R) = \frac{(0.0902)(0.0370)(0.0633)}{(0.0637)(0.0740)} = 0.0448
\]

(Numbers have been rounded to four decimal places for the sake of legibility. However, calculations were carried out with the unrounded numbers.)

Multiplying this by the benefit (in our example, $30,000) yields the fair value \(v\):

\[
v = (30,000)(0.0448) = \$1,343.89
\]

To cover costs apart from benefit payouts, we add $200, plus 10% of the fair value of the policy. This yields the premium \(r\):

\[
r = 1.1v + 200 = (1.1)(\$1,343.89) + \$200 = \$1,678.28
\]
4 The Excel file

Here, we explain the organization of the Excel file, and the formulas in various cells.

The user should enter the genre in cell C3, the region in cell C4, and the amount of the benefit in cell C5. These cells have been shaded in yellow. No other cells should be altered. The fair value of the policy should appear in cell C7, and the premium in cell C8.

**WARNING:** Misspelling the genre or the region will produce #DIV/0! errors in cells C7 and C8. A good way to avoid this is to go into the appropriate database, find an occurrence of your genre or region, and copy/paste it from the database into cell C7 or C8.

4.1 The databases

The worksheets “Data1” and “Data2” contain the database information.

In “Data1”, column A records the region in which each concert was scheduled; column B records “Yes” for cancelled concerts, and “No” for concerts that were not cancelled.

In “Data2”, column A records the genre of each concert; column B again includes “Yes” for cancelled concerts, and “No” for those that were not cancelled.

To prevent accidental alteration of the databases, the sheets “Data1” and “Data2” have been protected. No password is necessary to unprotect them.

The data-containing range “Data1!A1:B13865” has been given the name “data1”. The data-containing range “Data2!A1:B16707” has been given the name ”data2”. These names are used instead of the cell references in the calculations. If additional data is later added to one or both databases, the names “data1” and “data2” should be redefined to include the new data.

4.2 Calculations

The genre, region, and benefit are entered on the worksheet “Calculations”. The fair value of the policy and the premium appear on that sheet. The following table describes the contents of individual cells and ranges.

<table>
<thead>
<tr>
<th>Cell</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>Genre</td>
</tr>
<tr>
<td>C4</td>
<td>Region</td>
</tr>
<tr>
<td>C5</td>
<td>Benefit</td>
</tr>
</tbody>
</table>

These are the only cells that should be changed by the user.

<table>
<thead>
<tr>
<th>Cell</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>C7</td>
<td>Fair value = benefit · P(C</td>
</tr>
<tr>
<td>C8</td>
<td>Premium = 1.1 · (fair value) + $200</td>
</tr>
</tbody>
</table>

For the user’s convenience, these output values are placed directly below the input values. They are based on calculations in cells farther down the worksheet.
H10:I12 This range or parts of it are arguments to the DCOUNT function for “Data2”. H11 copies C4: it should not be changed by the user

H16:I18 This range or parts of it are arguments to the DCOUNT function for “Data1”. H17 copies C3: it should not be changed by the user.

B11 # of cancelled concerts in genre from “Data2”
B12 # of all cancelled concerts from “Data2”
B13 \( P(G|C) = \frac{B11}{B12} \)

B16 # of cancelled concerts in region from “Data1”
B17 # of all cancelled concerts from “Data1”
B18 \( P(R|C) = \frac{B16}{B17} \)

B21 # of cancelled concerts in “Data1”
B22 # of cancelled concerts in “Data2”
B23 # of all scheduled concerts in “Data1”
B24 # of all scheduled concerts in “Data2”
B25 # of all cancelled concerts in both databases: B21 + B22
B26 # of all scheduled concerts in both databases: B23 + B24
B27 \( P(C) = \frac{B25}{B26} \)

B30 # of all concerts in genre in “Data2”
B31 # of all scheduled concerts in “Data2”
B32 \( P(G) = \frac{B30}{B31} \)

B35 # of all concerts in region in “Data1”
B36 # of all scheduled concerts in “Data1”
B37 \( P(R) = \frac{B35}{B36} \)

B40 \( P(C|G \cap R) = \frac{P(G|C)P(R|C)P(C)}{P(G)P(R)} = \frac{B13 \cdot B18 \cdot B27}{B32 \cdot B37} \)