Sawtooth Software

RESEARCH PAPER SERIES

A Simple Introduction to TURF Analysis

John Howell
Sawtooth Software, Inc.

© Copyright 2010-2016, Sawtooth Software, Inc.
1457 E 840 N Orem, Utah
+1 801 477 4700
www.sawtoothsoftware.com

A Simple Introduction to TURF Analysis

By John Howell, Sawtooth Software, 2010
Updated 2012 & 2016

TURF is an acronym for Total Unduplicated Reach and Frequency. The purpose of TURF is to find a portfolio (or subset) of items that can be offered that will reach the widest audience possible. It does this by counting the number of respondents that would choose one of the items in the portfolio and how many of the items each respondent would choose from the portfolio. One of the places that TURF is traditionally used is to determine an advertising portfolio. Say you are an advertiser and are trying to decide which magazines to advertise in. You have a limited budget and can only advertise in a limited number of magazines, but you want to reach the widest audience possible. Since most people subscribe to more than one magazine, it would be ideal if you could place an ad in only one of the magazines that the person subscribed to and then use the rest of your budget to reach new people. That way you would have more people exposed to your ad, but keep the number of magazines that you needed to purchase ads from small.

An example may be helpful. You first need to find out which magazines a respondent subscribes to. Your survey could have a question like the following:

"Which of the following 6 magazines do you subscribe to?"

1. Wired
2. Popular Mechanics
3. Outdoor Life
4. Better Homes and Gardens
5. O
6. Newsweek

This would lead to 0 or 1 data with a 0 meaning the respondent does not subscribe and a 1 meaning that the respondent does subscribe.

The first step in TURF is to decide how big the portfolio is. For example, if I'm an advertiser, I may be able to advertise in only two magazines, so I would have a portfolio size of 2. The goal is to reach as many different people as possible, and this is done by not choosing magazines that would just reach the same audience. For example, there is probably a lot of overlap between Wired readers and Popular Mechanics readers so I wouldn't necessarily want to choose those as my only two magazines.

The first step is to define all n-way combinations of the magazines, where n is the number of items in the portfolio. In our example, there are 6 choose 2 combinations since we have 6 magazines and we want portfolios of 2 magazines.

The combinations are:

1. Wired and Popular Mechanics
2. Wired and Outdoor Life
3. Wired and Better Homes and Gardens
4. Wired and O
5. Wired and Newsweek
6. Popular Mechanics and Outdoor Life
7. Popular Mechanics and Better Homes and Gardens
8. Popular Mechanics and O
9. Popular Mechanics and Newsweek
10. Outdoor Life and Better Homes and Gardens
11. Outdoor Life and O
12. Outdoor Life and Newsweek
13. Better Homes and Garden and O
14. Better Homes and Gardens and Newsweek
15. O and Newsweek

Once you have set up the portfolios it is a simple matter to determine which respondents would read magazines in each portfolio. With TURF we can count not only Reach, but also Frequency.

If I have a respondent that subscribes to Wired, Outdoor Life, and Newsweek, then my table of reaches and frequencies would be:

Portfolios	Reach	Frequency
1. Wired and Popular Mechanics	1	1
2. Wired and Outdoor Life	1	2
3. Wired and Better Homes and Garden	1	1
4. Wired and O	1	1
5. Wired and Newsweek	1	2
6. Popular Mechanics and Outdoor Life	1	1
7. Popular Mechanics and Better Homes and Gardens	0	0
8. Popular Mechanics and O	0	0
9. Popular Mechanics and Newsweek	1	1
10. Outdoor Life and Better Homes and Gardens	1	1
11. Outdoor Life and O	1	1
12. Outdoor Life and Newsweek	1	2
13. Better Homes and Garden and O	0	0
14. Better Homes and Gardens and Newsweek	1	1
15. O and Newsweek	1	1

Reach is a 1 if any magazines that are subscribed to are included in the portfolio. Frequency is the number of magazines reaching the respondent in the portfolio.

If you add a second respondent that just subscribes to O then the updated table would be:

Portfolios	Reach	Frequency
1. Wired and Popular Mechanics	1	1
2. Wired and Outdoor Life	1	2
3. Wired and Better Homes and Garden	1	1
4. Wired and O	2	2
5. Wired and Newsweek	1	2
6. Popular Mechanics and Outdoor Life	1	1
7. Popular Mechanics and Better Homes and Gardens	0	0
8. Popular Mechanics and O	1	1
9. Popular Mechanics and Newsweek	1	1
10. Outdoor Life and Better Homes and Gardens	1	1
11. Outdoor Life and O	2	2
12. Outdoor Life and Newsweek	1	2
13. Better Homes and Garden and O	1	1
14. Better Homes and Gardens and Newsweek	1	1
15. O and Newsweek	2	2

In our two respondent case, advertising in Wired and O, Outdoor Life and O, or O and Newsweek would allow us to reach both respondents. The frequency column represents the number of times that a respondent would be exposed to the advertisement. For example, if we were to advertise in Wired and Outdoor Life, the first person would see the ad twice, but the second person would not see the ad at all. Thus, the reach is 1 person, but the frequency is two exposures.

MaxDiff TURF analysis has one additional wrinkle. MaxDiff scores are continuous values instead of discrete 0 or 1 values. Therefore, with MaxDiff we have one additional step; that is to "discretize" the data. We provide the option of doing that in two different ways and provide one additional method that is similar to TURF, but accounts directly for the MaxDiff scores.

The first method is to simply assume that respondents will only be reached if their first (top) choice is present in the portfolio. Therefore if a respondent had scores of:

1. Wired – 12.6
2. Popular Mechanics – 13.7
3. Outdoor Life – 6.3
4. Better Homes and Gardens – 31.2
5. O – 28.8
6. Newsweek – 7.4

Then only portfolios containing Better Homes and Gardens would count as reached. The problem with this is that it ignores a very good second choice. This reader would probably also be willing to read O magazine since it has a very close score to Better Homes and Gardens.

The second method of dealing with this is to set a threshold. If a respondent's score for an item is above the threshold then he/she is considered reached by that item. Using the same scores above, if the threshold were 25, then the respondent could be reached by portfolios that contained either O or Better Homes and Gardens. Therefore, it overcomes the first choice method's problem of good second choices being ignored. But, this method also has a small drawback in that you have to come up with a threshold value. This value is usually arbitrarily made up.

The third method that Sawtooth Software uses for MaxDiff data relies on the interpretation of the scores themselves. Using the logit formula on the raw MaxDiff scores (typically coming from HB estimation), we can calculate the probability of each item being chosen as best from a set of typical items like those shown in the in the MaxDiff questionnaire. We assume raw MaxDiff scores that are zero-centered, such that a typical (average) item has a score of 0.

To convert raw HB scores to probabilities of choice, we raise the constant e to the power of the utility—we exponentiate the utility, which in Excel is done using the formula =EXP(). It follows, then, that the likelihood of selecting item i from a set involving a-1 other items of average desirability is[1]:

$$P_i = e^{Ui}/(e^{Ui}+a-1)$$

It is easy to expand this equation to include more items in a portfolio. For example, the likelihood that a portfolio containing items i, j, and k reaches the respondent is:

$$P_{ijk} = (e^{Ui}+e^{Uj}+e^{Uk})/(e^{Ui}+e^{Uj}+e^{Uk}+a-1)$$

The result is a probability weighted reach where the probability of a respondent choosing the items is considered. We call this method of TURF *Weighted by Probability*. The advantage is that it takes into account strong second choices and does not require setting an arbitrary threshold.

When applying the Weighted by Probability method to anchored MaxDiff scores, the calculation of reach is only slightly different. Rather than compute the likelihood that the bundle of items would be picked instead of the a-1 other average items in the set, we compute reach as the likelihood that the bundle of items would be picked instead of the anchor item (which under anchored MaxDiff HB estimation is constrained to zero).

[1] If each of a items is assumed to have an average (zero) utility, raising e to the power of 0 yields 1 for each of those items. Thus, for a MaxDiff set containing a items, the exponentiated sum for the other a-1 average items in the set is equal to a-1.

It follows, then, that the likelihood of selecting item i instead of the anchor item (with utility equal to 0) is:

$$P_i = e^{Ui}/(e^{Ui}+1)$$

It is easy to expand this equation to include more items in a portfolio. For example, the likelihood that a portfolio containing items i, j, and k reaches the respondent (is chosen instead of the anchor item) is:

$$P_{ijk} = (e^{Ui}+e^{Uj}+e^{Uk})/(e^{Ui}+e^{Uj}+e^{Uk}+1)$$