Testing for significant changes in popularity

Denny H. Meyer

Department of Statistics, University of the Witwatersrand, Johannesburg, 2050 Republic of South Africa

A graphic procedure for determining whether a statistically significant change in television ratings has occurred is proposed. The method takes into account the fact that the ratings are generated from the same sample of viewers rather than from two independent samples. In addition it incorporates the effect of sample size. The method is applicable for the analysis of all market research data collected by such panel methods.

'n Prosedure om grafies te bepaal of 'n betekenisvolle verandering in die persentasie televisiekykers voorgekom het, word voorgestel. Die feit dat die persentasie kykers bepaal word uit dieselfde steekproef van kykers, in stede van uit twee onafhanklike steekproewe, word met dié metode in ag geneem. Daarbenewens word die effek van steekproefgrootte ook ingesluit. Die metode is van toepassing op die analise van alle marknavorsingsdata wat deur sulke 'paneel'-metodes ingesamel word.

Introduction

Television ratings are due to appear on the South African scene shortly. These ratings will be based on information collected from television monitors installed in 500 South African homes, which have been carefully chosen so as to be representative of South African television viewership. These ratings will be of considerable interest to advertisers, and, in particular, advertisers will want to test whether changes in ratings for particular programmes reflect a real change in popularity or whether the change is simply the result of sampling error. The statistical solution to this problem lends itself to a simple graphic analysis. The important parameters are found to be sample size, the (lower) rating, and the proportion of viewership common to both ratings.

Method

The solution to the above problem is based upon the McNemar test. This is a nonparametric test which can be viewed as a simple chi-squared goodness of fit test. The data are given in Table 1.

This table suggests a total sample of $N = A + B + C + D$ individuals. In the case of programme 1 the rating of viewing would have been $100(B + C)/N\%$ and in the case of programme 2 the rating of viewing would have been $100(A + C)/N\%$. We shall assume that the former rating is the lower. That is, programme 1 was less popular. In testing whether programme 1 was significantly less

Table 1 Viewership frequencies for two television programmes

	Programme 1	
Programme 2	Not viewing	Viewing
Not viewing	D	B
Viewing	A	C

Table 2 Chi-squared goodness of fit test

Change cells	Observed frequency	Expected frequency for no rating change
Viewed 1 but not 2	B	$(A + B)/2$
Viewed 2 but not 1	A	$(A + B)/2$

popular, attention focuses on the B and A cells in Table 1, because these are the only cells which reflect differences between viewership for the two programmes. If there is no significant difference in popularity between the two programmes then we would estimate the frequencies for these two cells to be $(A + B)/2$, giving the expected frequencies for the goodness of fit test in Table 2.

The chi-squared test statistic has one degree of freedom, a critical value of 3,841 when the probability of wrongly detecting a significant difference is set at a maximum of 5%, and can be simplified as shown below:

$$\Sigma \frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}} = \frac{(A-B)^2}{(A+B)}$$

If the difference between the two ratings,

$$100(A + C)/N\% - 100(B + C)/N\%$$

is called d, this means that a significant difference in ratings exists at a 5% significance level, whenever

$$\frac{(Nd)^2}{Nd+2B} > 3{,}841$$

i.e. $d > \dfrac{3{,}841 + \sqrt{3{,}841^2 + 30{,}728B}}{2N}$

Interpretation

The frequency B, which is the number of people who viewed programme 1 but not programme 2, is crucial in the above formula. If everyone who viewed programme 1 also viewed programme 2, then B would be zero, and a significant difference could be claimed whenever the rating difference exceeded $3,841/N$.

Alternatively, if none of the viewers of programme 1 viewed programme 2 then C would be zero. If k is the rating for programme 1, this means that a significant difference in viewership can be claimed in such a situation only when the rating difference exceeds the level:

$$\frac{3,841 + \sqrt{3,841^2 + 30,728Nk}}{2N}$$

The above cases, B equal to zero and C equal to zero represent the extremes. These two cases are shown in Figures 1 and 2 below for sample sizes of 400 and 600 respectively. Any difference in ratings less than $3,841/N$ can never be considered to be significant, even when everyone who watches programme 1 also watches programme 2. On the other hand, any rating difference of more than

$$\frac{3,841 + \sqrt{3,841^2 + 30,728Nk}}{2N}$$

will always be significant, even when nobody who watches programme 1 also watches programme 2.

Clearly, the larger the size of the random sample, the smaller the 'NEVER' area becomes. This means that increasing the sample size means that smaller rating differences will become significant. Similarly, increasing the sample size increases the area of the 'ALWAYS' region, meaning that a larger sample size reduces the rating difference that can be regarded as significant under all circumstances.

In practice, neither B nor C are zero. This means that significant rating differences will fall in the region

$$3,841/N; \quad \frac{3,841 + \sqrt{3,841^2 + 30,728Nk}}{2N}$$

In Figures 1 and 2 this region has been divided by a line portraying the case $B = Nk/2$. This case portrays the least significant difference which can be detected when half the people who viewed programme 1 also viewed programme 2. American studies (Nielsen) indicate that it is unusual for more than half the viewers to view the second programme if they viewed the first. Consequently, the area above this line denotes those
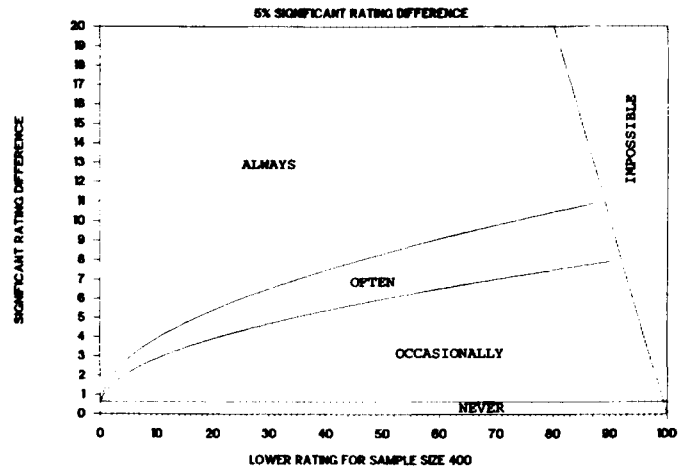

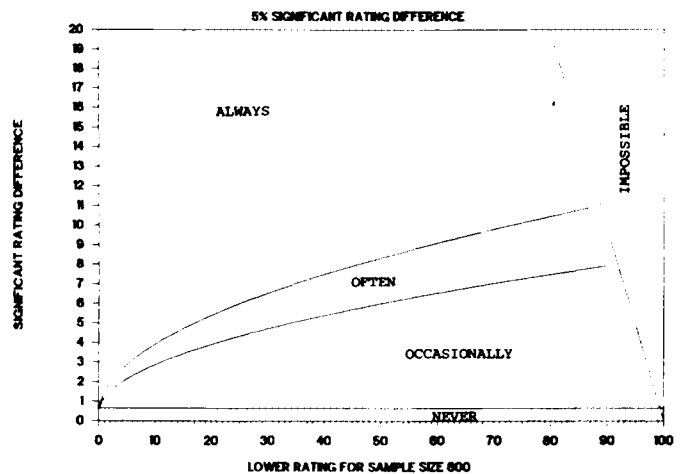Figure 1 5% Significant rating differences (N = 400)

Figure 2 5% Significant rating differences (N = 600)

differences which are often significant, whilst the area below this line indicates differences which are only occasionally significant.

Thus when the random sample sizes are 400 and 600, Figures 1 and 2 can be used to determine whether any given difference in television rating is likely to be statistically significant. If a rating difference falls in the 'OCCASIONALLY' region, an advertiser would be advised not to make any drastic changes to his advertising programme, but to monitor future performance carefully. If a rating difference falls in the 'OFTEN' region it can generally be safely assumed that a significant change is in progress. However, only when the rating difference falls in the 'ALWAYS' region can an advertiser be confident that a change has occurred.

Conclusion

The above graphic method has a lot of appeal for advertising managers who require an indication as to how wisely their television advertising budget is being spent, but do not want to get involved in statistical analysis. The method has many other possible applications. Most market research is practised on a

panel of individuals that remains relatively unchanged over time. It is incorrect to treat such repeat samples of data as being independent. A McNemar type of test, such as that described above, therefore has general applicability in the analysis of market research data.

Acknowledgement

References

Reference Supplement. 1986–87. Nielsen Station Index Methodology Techniques and Data Interpretation, p.53.

Siegel, S. 1956. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill. p.63–67.