

STATISTICAL INFERENCE ABOUT NET PROMOTER SCORES



15300 N. 90th Street • Suite #500 • Scottsdale, AZ 85260
+1.480.483.2700 • www.analyticalgroup.com

Dr. Albert Madansky
Vice President, The Analytical Group, Inc.
and
H.G.B. Alexander Professor Emeritus of Business
Administration Graduate School of Business
University of Chicago

The widely used Net Promoter Score (NPS) was introduced in Reichheld (2003), and is calculated based on responses to a single question: “How likely is it that you would recommend our company/product/service to a friend or colleague?” where the response is given based on a 0 to 10 scale. Respondents with scores of 9 or 10 are called “promoters,” those with scores of 6 or below are called “detractors,” and those with scores of 7 or 8 are called “passives.” The NPS is defined as 100 times the difference between the fraction of “promoters” and the fraction of “detractors” in the sample. There are a number of contexts in which researchers are interested in comparing NPSs to see if they are significantly different. In this note I will set forth the basic principles for performing such statistical comparisons.

1. Multinomial distribution basics

Consider a multinomial population with k categories whose corresponding means are $\theta_1, \dots, \theta_k$. Assume that we have a random sample of n observations from this population, with observed category proportions p_1, p_2, \dots, p_k , where

$$\sum_{i=1}^k p_i = 1$$

Since the sum of the p 's is 1, the p 's are negatively correlated (if one of the p_i is large the other p_j s will perform be small).

Suppose we wish to test the hypothesis that $\theta_1 = \theta_2$, based on the statistic $p_1 - p_2$. As shown by Wilks (1940), the standard deviation of the statistic $p_1 - p_2$ is

$$\sqrt{\frac{\theta_1(1-\theta_1) + \theta_2(1-\theta_2) + 2\theta_1\theta_2}{n}} = \sqrt{\frac{\theta_1 + \theta_2 - (\theta_1 - \theta_2)^2}{n}}$$

Since the θ 's are unknown, we estimate them by their sample estimates, and use as an estimate of the standard deviation of $p_1 - p_2$ the quantity

$$s = \sqrt{\frac{p_1 + p_2 - (p_1 - p_2)^2}{n}}$$

A common error made by those comparing two percentages from a multinomial distribution is to treat them as independent. In doing so they erroneously calculate the estimated standard deviation as

$$\sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{n}}$$

thereby underestimating the standard deviation and hence declaring non-significant differences as significant.

To apply this to NPS we let $k=3$, with subscript 1 representing “promoters,” subscript 2 representing “detractors,” and subscript 3 representing “passives.” Then (except for a factor of 100) $p_1 - p_2$ is the NPS and s given above is an estimate of the standard deviation of the NPS. If therefore one wanted to compare two NPSs, say L_1 and L_2 , computed from two independent samples of sizes n_1 and n_2 , respectively, one would calculate the z-statistic

$$z = \frac{L_1 - L_2}{\sqrt{\frac{p_{11} + p_{21} - (p_{11} - p_{21})^2}{n_1} + \frac{p_{12} + p_{22} - (p_{12} - p_{22})^2}{n_2}}}$$

where p_{11} and p_{12} are the “promoter” fractions and p_{21} and p_{22} are the “detractor” fractions in the respective independent samples.

Some researchers cleverly code the “promoters” as 1, the “detractors” as -1, and the “passives” as 0, calculate L_1 and L_2 as the averages of these coded data, and then use the two-sample t-test to test the significance of the difference. This leads to a minor error, which can be seen from the following analysis. One of the computations in facilitating the t-test is the calculation of the sample variance of L_a . Let the scores of the n_a respondents be denoted as y_1, \dots, y_{n_a} . Then the sample variance for sample a ($a = 1, 2$) is calculated as

$$s_a^2 = \frac{\sum_{i=1}^{n_a} y_i^2 - (\sum_{i=1}^{n_a} y_i)^2 / n_a}{n_a - 1} = \frac{n_a (p_{11} + p_{21}) - (n_a (p_{11} - p_{21}))^2 / n_a}{n_a - 1}$$

and the squared standard error of L_a is calculated as

$$\frac{s_a^2}{n_a} = \frac{p_{11} + p_{21} - (p_{11} - p_{21})^2}{n_a - 1}$$

instead of as

$$\frac{p_{11} + p_{21} - (p_{11} - p_{21})^2}{n_a}$$

While this error is negligible for large samples, it can make a difference for small samples.

2. Weighting

To develop the comparable procedure when the data are weighted, assume that the j -th member of the sample is given weight w_j , $j=1, \dots, n$. We represent the data as a collection of n_k 1's and 0's, with x_{ij} being a 1 if respondent j responded positively to category i and $x_{ij}=0$ otherwise. Then for example p_1 can be written as

$$p_1 = \frac{\sum_{j=1}^n x_{1j}}{n}$$

The weighted p_{w1} can be expressed as

$$p_{w1} = \frac{\sum_{j=1}^n w_j x_{1j}}{\sum_{j=1}^n w_j}$$

The determination of the standard deviation of the statistic $p_{w1} - p_{w2}$ is based on the following.

$$\begin{aligned} V(p_{w1} - p_{w2}) &= V\left[\frac{\sum_{j=1}^n w_j x_{1j}}{\sum_{j=1}^n w_j} - \frac{\sum_{j=1}^n w_j x_{2j}}{\sum_{j=1}^n w_j}\right] \\ &= \frac{\sum_{j=1}^n w_j^2 V(x_{1j} - x_{2j})}{\left(\sum_{j=1}^n w_j\right)^2} \end{aligned}$$

But

$$\begin{aligned} V(x_{1j} - x_{2j}) &= V(x_{1j}) + V(x_{2j}) - 2Cov(x_{1j}, x_{2j}) \\ &= \theta_1(1 - \theta_1) + \theta_2(1 - \theta_2) + 2\theta_1\theta_2 \end{aligned}$$

So

$$V(p_{w1} - p_{w2}) = \frac{\theta_1(1 - \theta_1) + \theta_2(1 - \theta_2) + 2\theta_1\theta_2}{f}$$

where f is the effective sample size (called the “design effect” by Kish (1965, Section 8.2, p. 258))

$$f = \frac{\sum_{j=1}^n w_j}{\sum_{j=1}^n w_j^2}$$

Since the θ 's are unknown, we estimate them by their sample estimate, and use as an estimate of the standard deviation of $p_{w1} - p_{w2}$ the quantity

$$s = \sqrt{\frac{p_1 + p_2 - (p_1 - p_2)^2}{f}}$$

If therefore one wanted to compare two NPSs, say L_{w1} and L_{w2} , computed from weighted data from two independent samples of effective sample sizes f_1 and f_2 , respectively, one would calculate the z-statistic

$$z = \frac{L_{w1} - L_{w2}}{\sqrt{\frac{p_{11} + p_{21} - (p_{11} - p_{21})^2}{f_1} + \frac{p_{12} + p_{22} - (p_{12} - p_{22})^2}{f_2}}}$$

3. More complex comparisons

Market researchers may sometimes compare NPSs from samples that may not be independent. One might ask a respondent the basic rating question about two different products. To complicate matters, some respondents may only rate one of the products rather than both. Obviously the ratings are correlated (because we may have the same respondent rating both products), so that the NPSs are correlated. To further complicate matters, the responses to each of the products rated by those respondents who rated both products may be assigned different weights.

Another scenario is one where the basic rating question is asked about a product, and then the sample is split into (not necessarily separate) subgroups, and one wants to know if the average NPSs are different for these subgroups. For example, we may ask whether the average NPS for users of product A is different from that of product B, where some of the respondents are users of both products. Again, to further complicate matters, the respondents may be weighted differently for product A and for product B.

REFERENCES

- Kish, Leslie 1965. Survey Sampling, New York: John Wiley & Sons.
- Reichheld, Frederick F. 2003. One Number You Need to Grow. Harvard Business Review 38:12 (December), 1-9.

Wilks, S.S. 1940. Confidence Limits and Critical Differences Between Proportions. Public Opinion Quarterly, 4:2 (June), 332-8.