Estudo do Erro Amostral aplicado ao NPS

1. Definições

Seja n o número de respondentes. Cada respondente pode ser detrator, neutro ou promotor, de acordo com a nota avaliada na pesquisa.

Seja:

#Promotores = P

#Neutros = N

#Detratores = D

#Total = T = n = P + N + D

Seja ainda p = P/n e d = D/n as proporções de promotores e detratores

Definição do NPS:

$$NPS = 100 * (p - d)$$

Pela propagação da variância da diferença, temos:

$$Var(NPS) = 100^2 * (Var(p) + Var(d) - 2 * Cov(p, d))$$

2. Revisão das distribuições de Bernoulli e Binomial

Para uma distribuição de **Bernoulli** discreta X, com probabilidade $P(X = 1) = p$ e $P(X = 0) = 1 - p$, tem-se que a esperança $E(X) = 0 * (1 - p) + 1 * p = p$ e a Variância dada por $V(X) = E\left(X^2\right) - E(X)^2 = 0^2 * (1 - p) + 1^2 * p - (E(X))^2 = p - p^2 = p(1 - p)$.

Retirado uma amostra aleatória sem reposição $X_1, X_2, \ldots, X_n$ de tamanho n e denotando por Y a quantidade de indivíduos igual a 1, tem-se $Y = X_1 + X_2 + \ldots + X_n$. Y tem distribuição **Binomial**.

$$Y = \sum_{i=1}^{n} X_i = n \sum_{i=1}^{n} \frac{X_i}{n} = n\overline{X}$$

Pelo teorema central do limite, $\overline{X}$ terá distribuição normal com média p e variância $\frac{p(1-p)}{n}$, enquanto que a transformação Y tem distribuição normal com média np e variância $np(1 - p)$

Pode-se observar que $\overline{X}$ é um próprio estimador de p, então pode-se considerar que a distribuição amostral de p é aproximadamente normal.

3. Obtenção do desvio padrão do NPS

Assumindo que a população N de possíveis respondentes é muito maior que o tamanho da amostra n, pode-se definir a Variância e a Covariância como:

$$Var(p) = \frac{p(1-p)}{n}$$

$$Cov(p, d) = -\frac{pd}{n}$$

Temos a expressão da variância e do desvio padrão:

$$Var(NPS) = 100^2 * \frac{p(1-p)+d(1-d)+2pd}{n}$$

$$\sigma(NPS) = 100 * \sqrt{\frac{p(1-p)+d(1-d)+2pd}{n}} \quad (1)$$

Para o caso onde não se sabe os valores da proporção de promotores e detratores, uma boa aproximação usada é a de que $p = 1/3$, $d = 1/3$. Var(NPS) = 2/3 = 0,6667. Além disso, para ser conservador, p=½ e d=½ fornece a maior Variância: Var(NPS)=1

Assim,

$$\sigma(NPS) = SE = 100 * \frac{0,8165}{\sqrt{n}} < 100 * \frac{1}{\sqrt{n}}$$

4. Obtenção do intervalo de confiança

O intervalo de confiança é dado por:

$$IC = [NPS - Z * \sigma; NPS + Z * \sigma]$$

Onde Z equivale ao número de desvios padrão do nível de confiança desejado.

Nível de Confiança	Z
99%	2,576
95%	1,96
90%	1,645
68,26%	1

Assim, para Z = 1,96 (95% de nível de confiança), pode-se obter o tamanho total do intervalo de confiança:

$$IC = 2 * Z * \sigma \approx \frac{320}{\sqrt{n}} < \frac{392}{\sqrt{n}}$$

$$n = \left(\frac{320}{IC}\right)^2 < \left(\frac{392}{IC}\right)^2$$

5. Quantidade de respostas para cada intervalo de confiança

Para um IC de $10\ (\pm 5)$, são necessárias 1537 respostas, por exemplo. Alguns valores são apresentados abaixo, para uma confiança de 95%.

IC NPS	Quantidade de Respostas (95%)
$20\ (\pm 10)$	385
$10\ (\pm 5)$	1540
$5\ (\pm 2,5)$	6150
$2\ (\pm 1)$	38.420
$1\ (\pm 0,5)$	154.000
$0,2\ (\pm 0,1)$	3.850.000

6. Quantidade de respostas x Quantidade de disparos

É importante ressaltar que os números apresentados estão relacionados a quantidade de respostas. Então, deve-se considerar ainda a taxa de resposta de seus clientes para obter a quantidade de disparos necessários. Usualmente esse valor se encontra entre 5% e 15%.

Isso indica que, para se obter um erro de 5 pontos percentuais no NPS é necessário de 7.000 a 20.000 disparos para se obter 1.000 avaliações.

7. Simplificando os cálculos

A expressão em (1) pode ser escrita de outra forma:

$$Var(NPS) = \tfrac{D}{T}\left(-1 - \tfrac{NPS}{100}\right)^2 + \tfrac{N}{T}\left(0 - \tfrac{NPS}{100}\right)^2 + \tfrac{P}{T}\left(1 - \tfrac{NPS}{100}\right)^2$$

Ou ainda

$$Var(NPS) = \left(\tfrac{P}{T} - \tfrac{D}{T}\right) - \left(\tfrac{P}{T} - \tfrac{D}{T}\right)^2$$

Já que, da definição de variância, $Var(X) = E\left((X-\mu)^2\right) = E\left(X^2\right) - E(X)^2$, respectivamente explicitados nas primeira e segunda fórmulas. A segunda é mais vantajosa do ponto de vista de cálculo e simplifica o entendimento, além de facilitar a escrita da fórmula do desvio padrão:

$$SE = \frac{\sqrt{(1-neutro)-(promotor-detrator)^2}}{\sqrt{avaliacoes\ respondidas}}$$

Em que ***neutro***, ***promotor*** e ***detrator*** se referem a porcentagem da amostra e ***população*** ao tamanho da amostra. Para se obter o erro no NPS, basta fazer 1,96 * SE, para um nível de confiança de 95%.

Por Marcos Gonçalves, Vinícius Pedroso e Henrique Rodrigues.