Incrementality Using Risk Ratio and Meta-analysis

Kiran Pannerselvam

Oct-Dec 2019

1 Introduction

If X is a random variable such that

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q = 1 - p, \end{cases} \tag{1}$$

then X follows a Bernoulli distribution.
If $Y = X_1 + ... + X_n$ and each of $X_1, ..., X_n$ has a Bernoulli distribution with the same probability of success p and they are independent, then the random variable for number of successes, Y has a binomial distribution, $B(n, p)$ with mean np and a standard deviation $\sqrt{npq}$, which can be approximated by

$$Z = \frac{Y - np}{\sqrt{npq}} \tag{2}$$

where Z is a standard normal random variable. This approximation holds good when $np > 5$ and $nq > 5$ (Read appendix A for further details.).
The proportion of success, $P = \frac{Y}{n}$ can also be approximated by a normal distribution, $\frac{Y}{n} = P \sim N(\frac{np}{n}, \frac{\sqrt{npq}}{n})$.

1.1 Percentage of increment between control conversion rate and test conversion rate

In this study, we consider two binomial populations, control-c and test-t. The control population refers to people to whom the ad is not shown whereas test population refers to people to whom the ad is shown. 'Success' is when an individual converts - chooses to buy the product - and we wish to draw inferences about proportions of conversions, p_c and p_t which are parameters of the two binomial populations. The proportion of conversions is same as conversion rate. We make use of the normal approximation described in the introduction to obtain confidence intervals. We construct two independent random samples of sizes n_c and n_t and the number of conversions (successes) in

Table 1: Contingency table

	Converted	Not Converted	Row Total
Control(C)	x_c	$n_c - x_c$	n_c
Test(T)	x_t	$n_t - x_t$	n_t
Column Total	$x_c + x_t$	$n_c - x_c + n_t - x_t$	$n_c + n_t$

the respective samples are measured as x_c and x_t then the sample proportions, $\hat{p_c} = \frac{x_c}{n_c}$ and $\hat{p_t} = \frac{x_t}{n_t}$ are assumed to have approximate normal distributions. Using proportions, three different measures of effect can be defined: odds ratio, risk ratio and risk difference. In our study, risk ratio (see Eq. (3)) is the appropriate measure because we are interested in measuring **incrementality** (see Eq. (4)). Table 1 is a 2x2 contingency table which can be used to represent the observations of a single study.

$$RR = \frac{p_c}{p_t} \tag{3}$$

$$inc = 1 - \frac{p_c}{p_t} \tag{4}$$

The logarithm of risk ratio whose distribution can be approximated by normal distribution is used for statistical analysis. As a result, the interval estimated will be in logarithmic scale which is converted to original scale using exponential. The logarithm of sample risk ratio Eq. (5) and variance Eq. (6) of the logarithm of sample risk ratio is estimated as

$$\hat{\theta} = \ln \frac{\hat{p_c}}{\hat{p_t}} \tag{5}$$

$$\hat{v} = \frac{1}{x_c} - \frac{1}{n_c} + \frac{1}{x_t} - \frac{1}{n_t} \tag{6}$$

The variance formula is obtained using delta method. The reader can refer to [1] for detailed derivation. Confidence interval (1-α)% for $\hat{\theta}$ is given as

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{\hat{v}}$$

using which the confidence interval is obtained as

$$CI = 1 - \exp\left(\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{\hat{v}}\right) \tag{7}$$

For a 95% confidence interval, $\alpha = 0.05$ and $z_{1-\frac{\alpha}{2}} = 1.96$. Equations (4) and (7) give the estimate of the effect (inc) and confidence interval for one study respectively.

Table 2: Contingency table with zero correction factor (= 0.5)

	Converted	Not Converted	Row Total
Control(C)	$x_c + 0.5$	$n_c - x_c + 0.5$	$n_c + 1$
Test(T)	$x_t + 0.5$	$n_t - x_t + 0.5$	$n_t + 1$
Column Total	$x_c + x_t + 1$	$n_c - x_c + n_t - x_t + 1$	$n_c + n_t + 2$

Zero correction: Some studies may report zero test or control conversions. In such cases, a zero correction factor is added to all the cells in the contingency table as shown in Table 2, to avoid numerical inconsistencies during statistical analysis. A constant correction factor of 0.5 is the norm, however a variable correction factor dependent on the population size of the particular group (control or test) can also be used. While a population dependent conversion factor is more balanced, it may lead to heavily inflated lower confidence bound.
An aggregate measure of the effect might be of interest when studies are conducted several times over a certain period of time or for a particular test case such as: *Campaign* or *Adset*. A meta-analysis over all the studies is performed to obtain the aggregate measure.

2 Meta-analysis

Meta-analysis is a statistical analysis that is used to summarize results from multiple studies measuring a particular effect. The aggregate estimate obtained as the end result of meta-analysis is more robust than the individual studies included because of the increased power. Each individual study considered in the meta-analysis comes with an estimate and confidence interval for risk ratio. While each study measures the same effect, there are differences in estimates. The source of this difference governs the type of model chosen for meta-analysis. There are two popular models for meta-analysis: fixed-effect model and random-effects model. Even though both models use similar formulae to obtain the aggregate estimate, they differ in their fundamental assumptions. In fixed-effect model, it is assumed that there is only one true effect underlying all the individual studies and the differences in the estimates is only due to sampling error. Whereas in random-effects model, the effects are allowed to vary from study to study because studies could vary based on type of sample chosen - age, gender and other demographics - or intensity of the treatment. Here, the effects from individual studies are assumed to be a random sample collected from the population of true effects. For further discussion about differences in fixed-effect and random-effects models, the reader is referred to [2]. Since such a variability in sampling and treatment is present in our ad-incrementality studies, a random-effects meta-analysis is the appropriate choice.

3 Steps involved in meta-analysis

Steps suggested in [4] are followed for performing meta-analysis of individual studies.

3.1 Weight of a study

If k studies need to be aggregated, equations (5) and (6) are used on the observed data (zero correction applied, if needed) to obtain logarithm of risk ratio ($\hat{\theta}_i$) and its variance ($\hat{v}_i$) for each study ($i = 1 : k$), and the inverse of the variance is used as the weight for each study.

$$w_i = \frac{1}{\hat{v}_i} \tag{8}$$

In a fixed-effect meta-analysis, these weights (w_i) along with $\hat{\theta}_i$ are used to calculate a weighted average which will be the aggregate measure and the inverse of the sum of weights will be the aggregate variance. However, in a random-effects model the variability of the effects (discussed in Section.2) needs to be considered for which there exist different heterogeneity measures.

3.2 Heterogeneity

Heterogeneity of the effects is measured using a moment estimator τ^2 proposed by [3].

$$\tau^2 = \begin{cases} \frac{Q-(k-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}} & \text{if } Q > k-1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where

$$\begin{aligned} Q &= \sum_{i=1}^{k} w_i \left(\hat{\theta}_i - \bar{\theta}\right)^2 \\ &= \sum_{i=1}^{k} w_i \left(\hat{\theta}_i - \left(\frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i}\right)\right)^2 \end{aligned} \tag{10}$$

simplifying the above expression, we get

$$Q = \sum_{i=1}^{k} w_i \hat{\theta}_i^2 - \frac{\left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right)^2}{\sum_{i=1}^{k} w_i}$$

As the formula suggests Q (popularly known as 'Cochran's Q') is the weighted sum of squared differences between individual studies and pooled study. Q follows a χ^2 distribution of $k-1$ degrees of freedom, where k is the number of studies considered in the meta-analysis.

3.3 Aggregate measure

Weights w_i are adjusted using τ^2 (see Eq. (11)) and the adjusted weights w_i^* are used to obtain the aggregate measure of the meta-analysis as per Eq. (12) and the aggregate variance is measured as per Eq. (13).

$$w_i^* = \frac{1}{v_i + \tau^2} \tag{11}$$

$$\theta^\dagger = \frac{\sum_{i=1}^{k} w_i^* \hat{\theta}_i}{\sum_{i=1}^{k} w_i^*} \tag{12}$$

$$v^\dagger = \frac{1}{\sum_{i=1}^{k} w_i^*} \tag{13}$$

The aggregate incrementality is calculated as $1 - exp(\theta^\dagger)$ and the confidence interval is obtained by replacing $\hat{\theta}$ and $\hat{v}$ in Eq. (7) with $\theta^\dagger$ and $v^\dagger$ respectively.

4 Meta-analysis of meta-analyses

Sometimes meta-analysis of meta-analyses needs to be performed, for instance, to obtain aggregate estimate of multiple *Adsets* of a company. Two methods are suggested in [5] based on availability of individual study information. First method uses all the individual studies of all *Adsets* and it follows the same procedure as [4] with one modification while calculating heterogeneity. Second method uses only the meta-analyses estimates. Consider J meta-analyses need to be aggregated each comprised of different number of individual studies k_j where $(j = 1 : J)$.

4.1 Method 1 - Using individual studies

The adjustment factor in Eq. (11) will instead be:

$$\tau_{m1}^2 = \begin{cases} \frac{Q_{m1} - \sum_{j=1}^{J}(k_j - 1)}{\sum_{i,j} w_{ij} - \frac{\sum_{i,j} w_{ij}^2}{\sum_{i,j} w_{ij}}} & \text{if } Q > \sum_{j=1}^{J}(k_j - 1) \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where

$$Q_{m1} = \sum_{i,j} w_{ij}\hat{\theta}_{ij}^2 - \frac{\left(\sum_{i,j} w_{ij}\hat{\theta}_{ij}\right)^2}{\sum_{i,j} w_{ij}} \tag{15}$$

Note that the only difference in the formulae is in calculating the degrees of freedom to compare against Q_{m1}. The weight for each individual study and the final aggregate and variance using method 1 is obtained as follows:

$$w_{ij}^* = \frac{1}{\hat{v}_{ij} + \tau_{m1}^2} \tag{16}$$

$$\theta_{m1}^\dagger = \frac{\sum_{i,j} w_{ij}^*\hat{\theta}_{ij}}{\sum_{i,j} w_{ij}^*} \tag{17}$$

$$v_{m1}^\dagger = \frac{1}{\sum_{i,j} w_{ij}^*} \tag{18}$$

4.2 Method 2 - Using meta-analyses

By aggregating J meta-analyses, the final aggregate and variance using method 2 is obtained as:

$$\theta_{m2}^\dagger = \frac{\sum_{j=1}^{J} w_j^*\hat{\theta}_j}{\sum_{j=1}^{J} w_j^*} \tag{19}$$

$$v_{m2}^\dagger = \frac{1}{\sum_{j=1}^{J} w_j^*} \tag{20}$$

where $\hat{\theta}_j$ and w_j^* are the effect estimate and variance of meta-analysis j, obtained by following procedure in Section.3. If a random-effects model is used during each of the J meta-analyses, the adjusted weight is obtained as per Eq. (21).

$$w_j^* = \frac{1}{\hat{v}_j + \tau_j^2} \tag{21}$$

where $\hat{v_j}$, the variance of the j^{th} meta-analysis and τ_j^2 and Q_j are obtained as per the procedure in Section.3.

$$\tau_j^2 = \frac{Q_j - (k_j - 1)}{\sum_{i=1}^{k_j} w_i - \frac{\sum_{i=1}^{k_j} w_i^2}{\sum_{i=1}^{k_j} w_i}} \tag{22}$$

$$Q_j = \sum_{i=1}^{k_j} w_i \hat{\theta}_i^2 - \frac{\left(\sum_{i=1}^{k_j} w_i \hat{\theta}_i\right)^2}{\sum_{i=1}^{k_j} w_i} \tag{23}$$

The difference between method 1 and 2 is in the calculation of the adjustment factor used for adjusting the weights which can be observed from equations (16) and (21). As the number of studies increases, the adjustment factors from both methods τ_{m1}^2 and τ_j^2 should get close to the true τ^2, subsequently the estimate of the effect and confidence interval from both methods will tend to be same.

A Expected frequency and normal distribution assumption

Expected frequencies are calculated assuming that the null hypothesis is true given the observed data i.e. the control conversion rate is same as test conversion rate.

Table 3: Contingency table with observed and expected frequencies

	Success (S)	Failure (F)	Row total
Control (C)	5 (2.9)	4518 (4520.1)	4523
Test (T)	1 (3.1)	4828 (4825.9)	4829
Column total	6	9346	9352

In Table 3, the proportion of success of the entire population is $\frac{6}{9352}$ which is obtained as shown in (Eq. (24)). Please note that we are not claiming that the proportion of success for the unobtainable population is $\frac{6}{9352}$. Given a population of 9352 where 6 is the total number of successes observed, if the null hypothesis was true then we'd expect same proportion of success in both control and test. Same logic is followed while calculating failure expectations as well. The expected frequency of success in control population is obtained as 2.9 against the observed frequency of 5, the detailed steps for which are shown below.

$$\begin{aligned} P(S) &= P(C) \times P(S|C) + P(T) \times P(S|T) \\ &= \frac{4523}{9352} \times \frac{5}{4523} + \frac{4829}{9352} \times \frac{1}{4829} \\ &= \frac{1}{9352} + \frac{5}{9352} \\ &= \frac{6}{9352} \end{aligned} \tag{24}$$

$$
\begin{aligned}
E(S \cap C) &= n \times P(S \cap C) \\
&= n \times P(S) \times P(C) \quad \text{(S and C are independent)} \\
&= 9352 \times \frac{6}{9352} \times \frac{4523}{9352} \\
&= 2.9
\end{aligned} \tag{25}
$$

Many texts on hypothesis testing say that the binomial distribution can be approximated by a normal distribution, only when the expected frequencies are greater than 5 in 80% of cells of the contingency table. However, this is just a rule of thumb and there is no general consensus on the minimum expected frequency, some authors have even suggested that it can be relaxed to fractional values. In any case, the simpler expression for calculating expected frequency is using row totals and column totals as shown in Eq. (26).

$$
E_{ij} = \frac{n_i n_j}{n} \tag{26}
$$

where E_{ij} is the expected frequency of population i in category j, n_i is the total number of observations in population i, n_j is the total number of observations in category j and $n = n_c + n_t$ is the total sample size. Expected frequency will be zero in only two possible scenarios: i. control/test population is zero, in which case the test was not conducted and ii. both control conversion and test conversion is zero, in which case, either we have too small a population to observe any conversions or that there is no conversion at all. We employ the zero correction factor in such cases but the confidence on the estimates will be poor and hence such studies are assigned less weight during aggregation.

References

[1] Stanislav Bashkyrtsev. *Relative Risk - Why doesn't standard error for ratios have log in it?* 2014. URL: https://stats.stackexchange.com/q/126727.

[2] Michael Borenstein et al. "A basic introduction to fixed-effect and random-effects models for meta-analysis". In: *Research Synthesis Methods* 1.2 (Apr. 2010), pp. 97–111. ISSN: 1759-2879. DOI: 10.1002/jrsm.12.

[3] Rebecca DerSimonian and Nan Laird. "Meta-analysis in clinical trials". In: *Controlled Clinical Trials* 7.3 (1986), pp. 177–188. ISSN: 01972456. DOI: 10.1016/0197-2456(86)90046-2.

[4] Jeruza L. Neyeloff, Sandra C. Fuchs, and Leila B. Moreira. "Meta-analyses and Forest plots using a microsoft excel spreadsheet: Step-by-step guide focusing on descriptive data analysis". In: *BMC Research Notes* 5 (2012). ISSN: 17560500. DOI: 10.1186/1756-0500-5-52.

[5] Liansheng Larry Tang, Michael Caudy, and Faye Taxman. "A statistical method for synthesizing meta-analyses". In: *Computational and Mathematical Methods in Medicine* 2013 (2013). ISSN: 1748670X. DOI: 10.1155/2013/732989.