Winter 2025 MATH 445 Project: Odds Ratio Confidence Intervals

Background and Summary

The purpose of this project is to conduct a simulation study to assess the reliability of the popular odds ratio (OR) confidence intervals in the 2 x 2 contingency table setting. Note that a 2 x 2 contingency table is given by:

	Success	Failure	Total
Group 1	n_{11}	n_{12}	n_{1+}
Group 2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	N

Also, their theoretical probabilities are given by:

	Success	Failure
Group 1	p_1	$1 - p_1$
Group 2	p_2	$1 - p_2$

As these theoretical probabilities p_1 and p_2 are unobtainable in practice, they are estimated by dividing each success count in the 2 x 2 contingency table by their respective row total. That is, $\hat{p}_1 = n_{11}/n_{1+}$ and $\hat{p}_2 = n_{21}/n_{2+}$. These estimates are called the maximum likelihood estimates (MLEs).

The OR is given by

$$\theta = \frac{p_1/(1 - p_1)}{p_2/(1 - p_2)}$$

and its MLE is given by

$$\hat{\theta} = \frac{\hat{p}_1/(1 - \hat{p}_1)}{\hat{p}_2/(1 - \hat{p}_2)} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

The OR confidence interval, suggested by Woolf (1955) based on the delta method, is a well-known confidence interval which is usually expressed in terms of the estimated log OR ($\log \hat{\theta}$) as follows:

$$\log \hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

To re-express this confidence interval in terms of the estimated OR ($\hat{\theta}$), we need to simply exponentiate the lower and upper bound of the confidence interval above. Lastly, when at least one of n_{ij} is zero, then the OR confidence interval is set to an entire nonnegative real line.

Unfortunately, Woolf's confidence interval above suffers from the fact that its actual coverage rate tends to be higher than the theoretical 95% when $\alpha = 0.05$ is chosen. Some adjustments have been suggested since then. A popular adjustment by Gart (1966) is given by:

$$\log \tilde{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{\tilde{n}_{11}} + \frac{1}{\tilde{n}_{12}} + \frac{1}{\tilde{n}_{21}} + \frac{1}{\tilde{n}_{22}}}$$

Here, $\tilde{n}_{ij} = n_{ij} + 0.5$, and $\tilde{\theta}$ is estimated using $\tilde{n}_{ij}$ instead of n_{ij}. After the lower and upper bound are calculated, they are exponentiated to obtain a confidence interval for OR.

Another modification, suggested by Agresti (1999), utilizes a similar adjustment as the one by Gart (1966) and can be expressed as follows:

$$\log \breve{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{\breve{n}_{11}} + \frac{1}{\breve{n}_{12}} + \frac{1}{\breve{n}_{21}} + \frac{1}{\breve{n}_{22}}}$$

Here, $\breve{n}_{ij} = n_{ij} + c_{ij}$, where $c_{ij} = 2n_{i+}n_{+j}/N^2$ is used, and $\breve{\theta}$ is estimated using $\breve{n}_{ij}$ instead of n_{ij}. Also, after the lower and upper bound are calculated, they are exponentiated to obtain a confidence interval for OR. Lastly, when at least one of n_{ij} is zero, then the OR confidence interval is set to an entire nonnegative real line.

A number of simulation studies have been published regarding the OR confidence intervals. Fagerland et al. (2011) note that these "approximate" OR confidence intervals tend to be conservative (i.e., have a higher coverage rate than the nominal coverage rate). Then, they recommend the Gart OR confidence interval as long as θ is somewhere in between 0.1 and 10. Lawson (2004) makes a similar recommendation, noting its satisfactory performance for small sample sizes. As a remark, note that the Gart OR confidence interval is also referred to as the Haldane-Anscombe correction or Gart adjusted logit confidence interval. Similarly, the Agresti OR confidence interval is also referred to as the independent-smoothed logit confidence interval.

A further adjustment to these three OR confidence intervals is possible. Recall that, in the two-sample t-test case, the most common version utilizes Welch's adjustment so that the common variance assumption can be dropped. A similar idea can be used for the OR confidence interval. By following the idea used for Welch's adjustment, for the Woolf OR interval, $z_{\alpha/2}$ may be replaced with $t_{\alpha/2,\nu}$, where:

$$\nu = \frac{2Nf_1^2}{f_3 - f_1^2}$$

and

$$f_j = \left(\frac{N}{n_{11}}\right)^j + \left(\frac{N}{n_{12}}\right)^j + \left(\frac{N}{n_{21}}\right)^j + \left(\frac{N}{n_{22}}\right)^j$$

In the case of the Gart and Agresti OR confidence intervals, n_{11}, n_{12}, n_{21}, n_{22}, and N in f_j are replaced with their respective adjusted cell counts.

Despite the popularity of Welch's adjustment for the two-sample t-test, its application to the OR confidence intervals has not been discussed in detail. Therefore, one of the main objectives of this project is to compare the performances of the three OR confidence intervals (Woolf, Gart, and Agresti) with those which utilize the Welch's adjustment.

Simulation Study Details

Simply explore different combinations of p_1, θ, n_{1+}, and n_{2+} to find situations in which these confidence intervals are reliable or unreliable, assuming a 95% confidence level (i.e., $\alpha = 0.05$). The R script file OR_CR.R shows how you can run simulations for certain combinations of p_1, θ, n_{1+}, and n_{2+}. Note that, as long as p_1 and θ are given, p_2 is automatically determined by

$$p_2 = \frac{p_1}{\theta(1 - p_1) + p_1}$$

Below are some things to consider when you run simulations:

- As mentioned above, both Fagerland et al. (2011) and Lawson (2004) recommend the Gart OR confidence interval based on its performance under small sample sizes and actual coverage rates. Is that always the case? In other words, can you find situations where the Gart OR confidence interval performs much worse than the other confidence intervals, and yet Welch's adjustment-based confidence intervals are still relatively reliable?
- To explore the point above in more detail, you should explore different sample sizes (can be made extremely small or extremely unequal, for example), parity of these sample sizes (odd-odd, odd-even, even-odd, and even-even), and different odds ratios (can be made extremely low or extremely high, e.g., way beyond the range [0.1, 10] recommended in Fagerland et al. (2011)). Note that, in the context of the simulation study, 0.1 is the same as 10 because the numerator and denominator of the odds ratio can be swapped.

Important: For this project, you are required to explore at least 1000 different combinations of these parameters. You can easily check how many combinations you have by looking at numsimcases in OR_CR.R. Also, each empirical coverage rate needs to be estimated using at least 10000 Monte Carlo simulations in the final report (i.e., set MC in the code to 10000 or higher).

Tips: Although you are required to do at least 10000 Monte Carlo simulations for each combination in the final report, to determine promising combinations of parameters, you should start with 1000 Monte Carlo simulations first to come up with your research plan. Once you determine combinations which are interesting enough, then you can increase the number of Monte Carlo simulations to 10000 or higher.

Presentation of the Results

One of the main challenges in this project is to come up with a convenient way to present your results. You may use tables to present the empirical coverage rates, but they are not necessarily helpful to get the 'big picture' of understanding when these confidence intervals are reliable.

If you see Figures 7 and 8 of Fagerland et al. (2011), they present their results using a plot with p_1 on the x-axis and the empirical coverage rates on the y-axis for different values of θ, n_{1+}, and n_{2+}. This would be an effective way to highlight the most important cases. By extending this, you could also try a 3D presentation of the results by looking at plotly (https://plotly.com/r/3d-charts/), for example. The package 'plotly' is easy to use, and it allows you to generate colorful 3D graphs which are visually appealing.

There may be other creative ways of presenting your results. As long as your results clearly show when these OR confidence intervals are reliable or unreliable, that is considered good enough.

Additional Details (Extra Credit, up to 6 points may be awarded)

You may choose to conduct extra analysis to enhance your research finding(s). Suggested topics are listed below:

- Present a real-life application to highlight advantages/disadvantages of certain OR confidence interval methods (highly recommended, up to 2 points).
- Conduct a simulation study to assess the cases where the confidence level is something other than 95% (up to 1 point).
- Perform an additional comprehensive literature search online (e.g., using Google Scholar) to see if your research finding(s) is also discussed in other papers (up to 1 point if additional paper(s) related to your finding(s) is found).
- Try resampling-based approach(es) (this would make the simulation study computationally expensive, so be careful) (up to 2 points).

Research Plan (Due Friday, February 28 at 11:59pm): 10 points

Before the deadline, talk to your classmates to see if they are interested in working together with you. Choose your group member(s) wisely and carefully. Each group consists of members of no more than four (two, three, or four), so choose your group member(s) carefully and wisely. If you would like to work alone, you need to obtain permission from the instructor ahead of time. Your request may be granted based on the homework performance and class attendance, unless you have other valid reason(s). Each group member will receive the same grade, so divide the work evenly taking into account each individual's strengths and weaknesses. Ideally, a group should consist of someone who is good at explaining mathematical ideas in words and someone who is good at programming in R.

Then, your group needs to submit a research plan as a PDF file before the deadline. Your research plan should contain the following:

- An informative title for your project (e.g., not just something like "Math 445 Project") and a list of group members. (**2** points)
- Your simulation study plan. This part needs to specify which combinations of p_1, θ, n_{1+}, and n_{2+} you are planning to use for your simulation study. You need to make sure that there are at least 1000 different combinations for your simulation study (**4** points).
- Rationale for your simulation plan. Discuss why you decided to choose the combinations above and what you are trying to achieve by examining these combinations. Also, you need to describe what your preliminary simulation results indicated (based on 1000 Monte Carlo simulations) (**4** points).

Note: The instructor only needs to see one submission per group. Please do not make multiple submissions from the same group.

Final Report (Due Friday, March 21 at 11:59m): 25 points

The final report guidelines closely follow that of the Undergraduate Research Project Competition. (USRESP: See https://www.causeweb.org/usproc/USRESP%20Competition for details.) The instructor may recommend the group(s) or individual(s) who came up with the best report to submit their work to either the USRESP Competition (Next Due Date: June 18, 2025; check the website) or other form(s) of public presentation. Whether or not your work will be selected heavily depends on the quality of writing, mathematical accuracy, and graphical and other forms of presentation of the results.

The final report guideline closely follows the USRESP Competition guidelines with some additional requirements. Specifically, your final report is a paper of no more than 20 pages reporting the results of your project including the following:

1. An up to three-page summary of background information in your own words. Here, you are asked to summarize background information and importance of the OR confidence intervals. It would be a good idea to include its brief history and applications to different fields of study (**4** points)

2. Simulation study (**10** points). This section is evaluated based on (i) whether you properly reported simulation results based on at least 1000 parameter combinations, (ii) the quality of your presentation of the results using tables, plots, and/or 3D plots, etc., and (iii) the accuracy of the interpretation of your simulation results. The combinations presented in the paper do not have to be the same as what you wrote in your research plan. However, please discuss with your instructor first in case you decide to significantly change your research plan. Also, you need to present the most important results, i.e., it is not necessary to present every single result in the report.
3. (Extra credit, up to 6 points) Optionally, present materials related to the extra credit mentioned above, such as real-life applications.
4. A discussion of the research, the limitations of the current research, reasonableness of any assumptions made, possibilities of future work/studies that should be conducted. (**4** points).

Note that you do not have to strictly follow the structure above. For example, you may combine some of these sections or re-order them. However, it is recommended to discuss with the instructor beforehand if you are planning to deviate from the suggested structure.

The entire written summary must be no more than 20 pages (single spaced, Arial 11-point font with standard 1-inch margins). If you are aiming for the USRESP competition, note that the judges may use the following criteria:

- Overall clarity and presentation
- Originality, creativity, and significance of the study
- Accuracy of simulation study, conclusions, and discussion

Also, refer to the reports from the past competition winners (https://www.causeweb.org/usproc/projects/winners) to understand how to write effectively.

In addition to the write-up of no more than 20 pages including the appendix, you should have

1. A separate title page which includes: The title of the project, names of the group members, and a one-paragraph abstract (summary) of the entire project with recommended length of no more than 150 words (**3** points).
2. References should be listed at the end of the report and do not count against the 20-page limit. If you did extra credit related to literature search, clearly indicate which references are related to it (**2** points).
3. The R code that can replicate all the results in the report. This will be submitted as a separate R script file. The R code must be well organized and documented (**2** points).

Only one submission per group is necessary. The submitter must submit the final report and R code (as an R script file) to Canvas. No late submission without valid excuse(s) will be accepted.

Word Processing Software

The most recommended word processing software is R Markdown or LaTeX simply because R Markdown or LaTeX let you easily present mathematical equations beautifully. You may choose to use Microsoft Word or Google Docs if you wish, but be careful when presenting your equations.

References

Agresti A. (1999). On logit confidence intervals for the odds ratio with small samples. *Biometrics* 55, 597—602.

Fagerland, M.W., Lydersen, S., and Laake, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical Methods in Medical Research* 24(2), 224—254.

Gart, J.J. (1966). Alternative analyses of contingency tables. *Journal of Royal Statistical Society, Series B: Statistical Methodology* 28, 164–179.

Lawson R. (2004). Small sample confidence intervals for the odds ratio. *Communications in Statistics: Simulation and Computation* 33(4), 1095—1113.

Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics* 19(4), 251—253.

(Note: Agresti (1999), Fagerland et al. (2015), and Lawson (2004) are included in the Project folder as well.)