**categorical data analysis from a dose response study**

# Abstract

Contents

[Abstract 1](#_Toc160553722)

[Introduction 3](#_Toc160553723)

[Data Summary 3](#_Toc160553724)

[Methods 4](#_Toc160553725)

[Results 5](#_Toc160553726)

# Introduction

This report presents a detailed dose-response analysis based on the data set provided at the end of this document. Dose-response analysis is an essential method in pharmacology and toxicology, offering insights into the relationship between a substance's dosage and its effect on the subjects under observation. Such analysis is critical for determining the optimal drug dosage that balances maximum efficacy with minimal side effects, a key step in developing and approving new medications.

The primary objective of this analysis is to explore the dose-response relationship across the varying dosage levels and to determine if there is a statistically significant trend indicating an increasing response with increasing doses. A critical part of this analysis is the implementation of a Multiple Comparison Procedure (MCP) to maintain control over the Type I error rate while conducting pairwise comparisons of each dose against the placebo. Moreover, this analysis aims to investigate the interaction effects between sex and treatment in these comparisons, employing the Breslow-Day test to examine the consistency of treatment effects across sex subgroups.

# Data Summary

The dataset presented for analysis offers a solid basis for examining the dose-response relationship within a clinical trial context. The study's design suggests it is a randomized controlled trial, characterized by several treatment levels, which facilitates a comparative analysis across different doses and a control group.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Response** | |  |  |
| **Treatment** | **1** | **0** | **Total** | **Proportion (%)** |
| 0 | 11 | 45 | 56 | 25.81 |
| 1 | 21 | 33 | 54 | 24.88 |
| 2 | 24 | 30 | 54 | 24.88 |
| 3 | 28 | 25 | 53 | 24.42 |
| **Total** | 84 | 133 | 217 | 100.00 |

Table 1. Distribution of Treatment Responses and Proportions by Dose Level

The dataset comprises four treatment levels, where Dose 0 serves as the control group, with subjects receiving a placebo. This group is essential for establishing a baseline response level and assessing the placebo effect—a phenomenon where subjects exhibit a response even without receiving an active drug component. Doses 1 through 3 represent increasing concentrations of drug, aimed at evaluating whether higher doses correspond to an increased likelihood of response. These varying doses—low, medium, and high—are intended to assess the drug's efficacy and safety across different concentrations.

The dataset's primary endpoint is the binary response variable 'resp', indicating whether a subject is a responder (resp=1) or a non-responder (resp=0). A key part of the analysis will focus on the number of responders in each treatment group. Typically, one would expect to observe a dose-dependent increase in responders, indicating a positive dose-response relationship.

This dataset also includes a variable labeled 'ctr', which allows for the aggregation and comparison of data from different centers. Additionally, 'pid' represents the patient identification number, a unique alphanumeric code assigned to each study participant. Such the analysis could evaluate the drug's effects across different environments and individual characteristics.

# Methods

Multiple Comparison Procedures (MCP) are a collection of statistical methodologies designed to handle the challenges associated with making inferences when multiple hypotheses are being tested simultaneously. In the context of dose-response studies, or any scientific investigation involving several treatment groups, MCPs are particularly critical as they help to control for Type I error, commonly known as a "false positive.". When multiple pairwise comparisons are made, the probability of encountering at least one Type I error increases with the number of comparisons. To address this issue, MCPs adjust the significance level, thereby controlling the probability of making one or more Type I errors across the entire family of tests.

There are some well known MCPs. In this report, we use Bonferroni correction. The Bonferroni correction is a multiple comparisons correction method used in statistical hypothesis testing. It is designed to counteract the problem of Type I errors that occur when multiple pairwise tests are conducted simultaneously. In this report, we will perform multiple statistical tests on efficacy of each dose level to the placebo. In Bonferroni correction, we need to Determine the number of comparisons. In this report, we will have three comparisons when each treatment is compared to the placebo. Then, we take the desired overall alpha level () and divide it by the number of comparisons. Any individual test with a p-value below adjusted significant level () would be considered significant after the Bonferroni correction.

The Cochran-Mantel-Haenszel (CMH) test is a statistical analysis method used primarily in epidemiological studies to evaluate the association between an exposure and an outcome, while controlling for one or more confounding variables. It provides a way to test for an overall association between the exposure and the outcome across all strata. CMH test is particularly useful when dealing with categorical data organized into several strata, with each stratum being a 2x2 contingency table that represents different levels of a confounding variable.

The Cochran-Armitage Trend Test is a versatile and widely used method in the analysis of categorical data, particularly in the context of dose-response studies, allowing researchers to make more informed decisions about the efficacy and safety of treatments. Developed by William G. Cochran and Peter Armitage, Cochran-Armitage Trend Test is a modification of the Chi-square test. It is a statistical method used primarily to analyze categorical data when the independent variable is ordinal, and the dependent variable is binary. This type of test is particularly useful in dose-response studies to determine if there is a consistent trend in the proportion of response as the dose increases.

An integral part of the analysis involves exploring potential interactions between sex and treatment across each pairwise comparison. In this report, we will use the Breslow-Day test to assess the homogeneity of the odds ratios across different sex. The Breslow-Day Test, named after statisticians Norman E. Breslow and Nicholas E. Day, is a statistical test used to verify the homogeneity of the odds ratios across different strata. This test is specifically designed to assess whether the association between an exposure and an outcome is consistent across various levels of a third variable. The primary use of the Breslow-Day Test is to evaluate the hypothesis that the odds ratios from different strata are the same. Breslow-Day test provides a way to ensure that the estimate of the odds ratio across strata is meaningful and that it does not mask important subgroup differences.

# Results

## 4.1 Association Tests

In this report, we use CMH test to test the association between each dose and placebo. The hypotheses for CMH test are

* Null Hypothesis (): There is no association between the dose and the responder
* Alternative Hypothesis (): There is an association between the dose and the responder