

TYPE I ERROR ASSESSMENT AND POWER COMPARISON OF ANOVA AND ZERO-
INFLATED METHODS ON ZERO-INFLATED DATA

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Lucas Blackmore Young

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

April 2019

Fargo, North Dakota

North Dakota State University
Graduate School

Title

TYPE I ERROR ASSESSMENT AND POWER COMPARISON OF
ANOVA AND ZERO-INFLATED METHODS ON ZERO-INFLATED
DATA

By

Lucas Blackmore Young

The Supervisory Committee certifies that this *disquisition* complies with North Dakota
State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel

Chair

Curt Doetkott

Dr. Glenn Dorsam

Approved:

3 May 2019

Date

Dr. Rhonda Magel

Department Chair

ABSTRACT

Many tests for the analysis of continuous data have the underlying assumption that the data in question follows a normal distribution (ex. ANOVA, regression, etc.). Within certain research topics, it is common to end up with a dataset that has a disproportionately high number of zero-values but is otherwise relatively normal. These datasets are often referred to as ‘zero-inflated’ and their analysis can be challenging. An example of where these zero-inflated datasets arise is in plant science. We conducted a simulation study to compare the performance of zero-inflated models to a standard ANOVA model on different types of zero-inflated data. Underlying distributions, experimental design scenario, sample sizes, and percentages of zeros were variables of consideration. In this study, we conduct a Type I error assessment followed by a power comparison between the models.

TABLE OF CONTENTS

| | |
|---|-----|
| ABSTRACT..... | iii |
| LIST OF TABLES..... | vi |
| LIST OF FIGURES..... | vii |
| LIST OF APPENDIX TABLES..... | ix |
| LIST OF APPENDIX FIGURES..... | x |
| CHAPTER 1. INTRODUCTION..... | 1 |
| CHAPTER 2. LITERATURE REVIEW..... | 4 |
| Zero-Inflated Data..... | 4 |
| ANOVA..... | 5 |
| Zero-Inflated Methods..... | 5 |
| CHAPTER 3. METHODS..... | 8 |
| Simulation Study Overview..... | 8 |
| Simulation Parameters..... | 10 |
| Type I Error Assessment..... | 13 |
| Realized Effect Size..... | 14 |
| Power Comparison..... | 15 |
| Simulation Study Outline..... | 17 |
| SAS Code..... | 18 |
| CHAPTER 4. RESULTS..... | 19 |
| Two-Population Type I Error Assessment..... | 19 |
| Eight-Population Type I Error Assessment..... | 23 |
| Zero-Inflated Normal Power Comparison..... | 25 |
| Zero-Inflated Poisson Power Comparison..... | 32 |
| Zero-Inflated Negative Binomial Power Comparison..... | 39 |

| | |
|---|-----|
| Research Questions | 46 |
| CHAPTER 5. CONCLUSION..... | 48 |
| Future Considerations | 50 |
| REFERENCES | 52 |
| APPENDIX A. TYPE I ERROR ASSESSMENT RESULTS | 54 |
| Two-Population Normal Samples | 54 |
| Two-Population Poisson Samples..... | 56 |
| Two-Population Negative Binomial Samples | 58 |
| Eight-Population Normal Samples..... | 60 |
| Eight-Population Poisson Samples..... | 61 |
| Eight-Population Negative Binomial Samples..... | 62 |
| APPENDIX B. POWER COMPARISON RESULTS..... | 63 |
| Two-Population Scenario | 63 |
| Eight-Population 1 vs 7 Scenario | 81 |
| Eight-Population 4 vs 4 Scenario | 99 |
| APPENDIX C. EIGHT-POPULATION SAS CODE | 117 |
| Normal Samples | 117 |
| Poisson Samples | 122 |
| Negative Binomial Samples | 127 |

LIST OF TABLES

| <u>Table</u> | <u>Page</u> |
|--|-------------|
| 1. Realized Effect Size..... | 15 |
| 2. Normal Sample – Normal Model Two-Population Type I Error..... | 20 |
| 3. Normal Sample – Poisson Model Two-Population Type I Error. | 20 |
| 4. Normal Sample – ZIP Model Two-Population Type I Error..... | 20 |
| 5. Poisson Sample – Normal Model Two-Population Type I Error. | 21 |
| 6. Poisson Sample – Poisson Model Two-Population Type I Error. | 21 |
| 7. Poisson Sample – ZIP Model Two-Population Type I Error..... | 21 |
| 8. Negative Binomial Sample – Normal Model Two-Population Type I Error..... | 22 |
| 9. Negative Binomial Sample – Poisson Model Two-Population Type I Error. | 22 |
| 10. Negative Binomial Sample – ZIP Model Two-Population Type I Error..... | 22 |
| 11. Normal Sample – Normal Model Eight-Population Type I Error. | 23 |
| 12. Normal Sample – Poisson Model Eight-Population Type I Error. | 23 |
| 13. Normal Sample – ZIP Model Eight-Population Type I Error..... | 23 |
| 14. Poisson Sample – Normal Model Eight-Population Type I Error. | 24 |
| 15. Poisson Sample – Poisson Model Eight-Population Type I Error..... | 24 |
| 16. Poisson Sample – ZIP Model Eight-Population Type I Error. | 24 |
| 17. Negative Binomial Sample – Normal Model Eight-Population Type I Error. | 25 |
| 18. Negative Binomial Sample – Poisson Model Eight-Population Type I Error. | 25 |
| 19. Negative Binomial Sample – ZIP Model Eight-Population Type I Error. | 25 |

LIST OF FIGURES

| <u>Figure</u> | <u>Page</u> |
|--|-------------|
| 1. Normal Distribution – $\mu = 20, \sigma = 2, p = 0.1$ | 12 |
| 2. Poisson Distribution – $\lambda = 20, p = 0.1$ | 12 |
| 3. Negative Binomial Distribution – $k(nb) = 20, p(nb) = 0.5, p = 0.1$ | 13 |
| 4. Normal Sample – $p = 0.0$ Two-Population Power (1σ). | 26 |
| 5. Normal Sample – $p = 0.1$ Two-Population Power (1σ). | 27 |
| 6. Normal Sample – $p = 0.2$ Two-Population Power (1σ). | 27 |
| 7. Normal Sample – $p = 0.3$ Two-Population Power (1σ). | 28 |
| 8. Normal Sample – $p = 0.4$ Two-Population Power (1σ). | 28 |
| 9. Normal Sample – $p = 0.5$ Two-Population Power (1σ). | 29 |
| 10. Normal Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ). | 29 |
| 11. Normal Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1σ). | 30 |
| 12. Normal Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1σ). | 30 |
| 13. Normal Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1σ). | 31 |
| 14. Normal Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1σ). | 31 |
| 15. Normal Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ). | 32 |
| 16. Poisson Sample – $p = 0.0$ Two-Population Power (1σ). | 33 |
| 17. Poisson Sample – $p = 0.1$ Two-Population Power (1σ). | 34 |
| 18. Poisson Sample – $p = 0.2$ Two-Population Power (1σ). | 34 |
| 19. Poisson Sample – $p = 0.3$ Two-Population Power (1σ). | 35 |
| 20. Poisson Sample – $p = 0.4$ Two-Population Power (1σ). | 35 |
| 21. Poisson Sample – $p = 0.5$ Two-Population Power (1σ). | 36 |
| 22. Poisson Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ). | 36 |
| 23. Poisson Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1σ). | 37 |

| | |
|---|----|
| 24. Poisson Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1σ)..... | 37 |
| 25. Poisson Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1σ)..... | 38 |
| 26. Poisson Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1σ)..... | 38 |
| 27. Poisson Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ)..... | 39 |
| 28. Negative Binomial Sample – $p = 0.0$ Two-Population Power (1σ)..... | 40 |
| 29. Negative Binomial Sample – $p = 0.1$ Two-Population Power (1σ)..... | 41 |
| 30. Negative Binomial Sample – $p = 0.2$ Two-Population Power (1σ)..... | 41 |
| 31. Negative Binomial Sample – $p = 0.3$ Two-Population Power (1σ)..... | 42 |
| 32. Negative Binomial Sample – $p = 0.4$ Two-Population Power (1σ)..... | 42 |
| 33. Negative Binomial Sample – $p = 0.5$ Two-Population Power (1σ)..... | 43 |
| 34. Negative Binomial Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ)..... | 43 |
| 35. Negative Binomial Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1σ)..... | 44 |
| 36. Negative Binomial Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1σ)..... | 44 |
| 37. Negative Binomial Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1σ)..... | 45 |
| 38. Negative Binomial Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1σ)..... | 45 |
| 39. Negative Binomial Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ)..... | 46 |

LIST OF APPENDIX TABLES

| <u>Table</u> | <u>Page</u> |
|---|-------------|
| A1. Normal Sample – Normal Model Two-Population Type I Error..... | 54 |
| A2. Normal Sample – Poisson Model Two-Population Type I Error..... | 54 |
| A3. Normal Sample – ZIP Model Two-Population Type I Error. | 55 |
| A4. Poisson Sample – Normal Model Two-Population Type I Error..... | 56 |
| A5. Poisson Sample – Poisson Model Two-Population Type I Error. | 56 |
| A6. Poisson Sample – ZIP Model Two-Population Type I Error..... | 57 |
| A7. Negative Binomial Sample – Normal Model Two-Population Type I Error..... | 58 |
| A8. Negative Binomial Sample – Poisson Model Two-Population Type I Error. | 58 |
| A9. Negative Binomial Sample – ZIP Model Two-Population Type I Error. | 59 |
| A10. Normal Sample – Normal Model Eight-Population Type I Error. | 60 |
| A11. Normal Sample – Poisson Model Eight-Population Type I Error. | 60 |
| A12. Normal Sample – ZIP Model Eight-Population Type I Error..... | 60 |
| A13. Poisson Sample – Normal Model Eight-Population Type I Error. | 61 |
| A14. Poisson Sample – Poisson Model Eight-Population Type I Error..... | 61 |
| A15. Poisson Sample – ZIP Model Eight-Population Type I Error. | 61 |
| A16. Negative Binomial Sample – Normal Model Eight-Population Type I Error. | 62 |
| A17. Negative Binomial Sample – Poisson Model Eight-Population Type I Error. | 62 |
| A18. Negative Binomial Sample – ZIP Model Eight-Population Type I Error..... | 62 |

LIST OF APPENDIX FIGURES

| <u>Figure</u> | <u>Page</u> |
|---|-------------|
| B1. Normal Sample – Normal Model Two-Population Power (1σ)..... | 63 |
| B2. Normal Sample – Poisson Model Two-Population Power (1σ)..... | 64 |
| B3. Normal Sample – ZIP Model Two-Population Power (1σ)..... | 65 |
| B4. Poisson Sample – Normal Model Two-Population Power (1σ)..... | 66 |
| B5. Poisson Sample – Poisson Model Two-Population Power (1σ)..... | 67 |
| B6. Poisson Sample – ZIP Model Two-Population Power (1σ)..... | 68 |
| B7. Negative Binomial Sample – Normal Model Two-Population Power (1σ)..... | 69 |
| B8. Negative Binomial Sample – Poisson Model Two-Population Power (1σ)..... | 70 |
| B9. Negative Binomial Sample – ZIP Model Two-Population Power (1σ)..... | 71 |
| B10. Normal Sample – Normal Model Two-Population Power (2σ)..... | 72 |
| B11. Normal Sample – Poisson Model Two-Population Power (2σ)..... | 73 |
| B12. Normal Sample – ZIP Model Two-Population Power (2σ)..... | 74 |
| B13. Poisson Sample – Normal Model Two-Population Power (2σ)..... | 75 |
| B14. Poisson Sample – Poisson Model Two-Population Power (2σ)..... | 76 |
| B15. Poisson Sample – ZIP Model Two-Population Power (2σ)..... | 77 |
| B16. Negative Binomial Sample – Normal Model Two-Population Power (2σ)..... | 78 |
| B17. Negative Binomial Sample – Poisson Model Two-Population Power (2σ)..... | 79 |
| B18. Negative Binomial Sample – ZIP Model Two-Population Power (2σ)..... | 80 |
| B19. Normal Sample – Normal Model Eight-Population Power (1 vs 7, 1σ)..... | 81 |
| B20. Normal Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ)..... | 82 |
| B21. Normal Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ)..... | 83 |
| B22. Poisson Sample – Normal Model Eight-Population Power (1 vs 7, 1σ)..... | 84 |
| B23. Poisson Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ)..... | 85 |

| | |
|--|-----|
| B24. Poisson Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ). | 86 |
| B25. Negative Binomial Sample – Normal Model Eight-Population Power (1 vs 7, 1σ). | 87 |
| B26. Negative Binomial Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ). | 88 |
| B27. Negative Binomial Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ). | 89 |
| B28. Normal Sample – Normal Model Eight-Population Power (1 vs 7, 2σ). | 90 |
| B29. Normal Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ). | 91 |
| B30. Normal Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ). | 92 |
| B31. Poisson Sample – Normal Model Eight-Population Power (1 vs 7, 2σ). | 93 |
| B32. Poisson Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ). | 94 |
| B33. Poisson Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ). | 95 |
| B34. Negative Binomial Sample – Normal Model Eight-Population Power (1 vs 7, 2σ). | 96 |
| B35. Negative Binomial Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ). | 97 |
| B36. Negative Binomial Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ). | 98 |
| B37. Normal Sample – Normal Model Eight-Population Power (4 vs 4, 1σ). | 99 |
| B38. Normal Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ). | 100 |
| B39. Normal Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ). | 101 |
| B40. Poisson Sample – Normal Model Eight-Population Power (4 vs 4, 1σ). | 102 |
| B41. Poisson Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ). | 103 |
| B42. Poisson Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ). | 104 |
| B43. Negative Binomial Sample – Normal Model Eight-Population Power (4 vs 4, 1σ). | 105 |
| B44. Negative Binomial Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ). | 106 |
| B45. Negative Binomial Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ). | 107 |
| B46. Normal Sample – Normal Model Eight-Population Power (4 vs 4, 2σ). | 108 |
| B47. Normal Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ). | 109 |
| B48. Normal Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ). | 110 |

| | |
|--|-----|
| B49. Poisson Sample – Normal Model Eight-Population Power (4 vs 4, 2σ)..... | 111 |
| B50. Poisson Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ)..... | 112 |
| B51. Poisson Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ)..... | 113 |
| B52. Negative Binomial Sample – Normal Model Eight-Population Power (4 vs 4, 2σ)..... | 114 |
| B53. Negative Binomial Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ)..... | 115 |
| B54. Negative Binomial Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ)..... | 116 |

CHAPTER 1. INTRODUCTION

Many tests for the analysis of continuous data have the underlying assumption that the samples under consideration come from a normal distribution (ex. ANOVA, regression, etc.). In some fields of research, it is common to end up with samples having a disproportionately high number of zero-values but appear to come from populations that are approximately normally distributed. These samples are often referred to as ‘zero-inflated’ (Min & Agresti, 2002), and their analysis can be challenging. Because of the high number of zeros, the distributions from which the samples are drawn are not normal. Thus, analysis techniques with the underlying assumption of normality can become inaccurate and misleading. It is generally understood that most statistical models can handle small deviations from the normal distribution assumption and still be accurate, but of the effect of zero-inflation on test results has not been quantified.

An example of where these zero-inflated datasets arise is in plant science. Suppose an experiment is designed where treatments are applied to seeds or their soil, and the height of the emerging plants is measured as the dependent variable. Generally, some percentage of the seeds do not grow; this results in a zero-inflated height dataset. Other sources of zero-inflated datasets are research on cold spells, defects in manufacturing, household expenditures on durable goods, alcohol consumption, insurance benefits, healthcare expenditures, and motor vehicle crash records (Eggers, 2015; Lambert, 1992; Lord et al., 2005).

Analysis of variance (ANOVA) is a statistical method that can be used to test if population means are equal using samples from the populations in question (Fisher, 1963). Within the context of testing for equality of means, ANOVA is an extension of the two-sample t-test that can be used for more than two samples. Both ANOVA and the two-sample t-test share the assumptions of sample independence, samples having an underlying normal distribution, and

homogeneity of variance. The potential problem with using ANOVA on zero-inflated normal data is that the zero-inflation violates the assumption that the samples have an underlying normal distribution. Zero-inflation can also lead to a violation in the homogeneity of variance assumption.

Models designed for dealing with zero-inflated data exist, but their performance compared to ANOVA has not been quantified. Examples of zero-inflated models include hurdle, Tobit, and zero-inflated Poisson (Stroup, 2013; Tobin, 1958; Greene, 1994). These models utilize different means to differentiate between zero and non-zero portions of the model to account for the extra zero values. With these different methods comes the potential for optimization on which methods are most powerful in a given set of circumstances. It is also possible that in some situations the zero-inflated models don't provide any advantage over standard ANOVA, and subsequently aren't worth dealing with the added complexity.

To compare ANOVA and zero-inflated models, we will conduct two phases of analysis. The first will be a Type I error assessment, and the second will be a power comparison. For the Type I error assessment, we will sample different distributions with varying levels of zero-inflation and sample size. For each combination of underlying distribution, level of zero-inflation, and sample size, we will fit each of the models and calculate the Type I error rate across many samples. After observing the ability of the models to control Type I error rate, we will proceed with a power comparison. In this phase, each distribution will again be sampled with varying levels of zero-inflation and sample size. We again fit the models and calculate the power of each across many samples. To conduct the various phases of this study, we will select simulation parameters by analyzing samples from an existing study from plant science. The parameters we find in the plant science study will be used for our simulations.

In addition to the impacts of the underlying distribution, the level of zero-inflation, and same size, we will also explore the effect of some different experimental design scenarios. We will initially consider a two-population scenario with a baseline and a treatment population. We will then expand the scope to two different eight-population cases: one with seven baseline populations and a single treatment population, and another with four baseline populations and four treatment populations. With the results from these different experimental design scenarios, we can evaluate the performance of ANOVA and zero-inflated models within different experimental contexts.

In this study we will also seek to answer the following research questions:

- 1) How does Type I error compare between models specifically designed for zero-inflated data and standard ANOVA?
- 2) Do models specifically designed for zero-inflated data exhibit improved power over ANOVA across different sample sizes?
- 3) What level of zero-inflation significantly affects the Type I error and power of ANOVA?

In the following sections we will provide additional information on the current state of zero-inflated data analysis and review past research on the topic. Next, we will outline the methodology used to conduct this study which will include a defined scope for the simulation parameters. Following the methodology section, we will present simulation results and provide discussion around them. Finally, we will summarize this study in the conclusion.

CHAPTER 2. LITERATURE REVIEW

In this section we will provide additional information on the current state of zero-inflated data analysis and present past research on the topic. This will include discussion around the sources of zero-inflated data, the problems associated with using ANOVA on zero-inflated data, and proposed alternative solutions for analyzing zero-inflated data.

Zero-Inflated Data

In addition to the zero-inflated datasets that can arise in plant sciences, many other areas of study can produce this type of data. Research on cold spells, defects in manufacturing, household expenditures on durable goods, alcohol consumption, insurance benefits, and healthcare expenditures are a few of these areas (Eggers, 2015; Lambert, 1992). Regarding defects in manufacturing, when a machine is properly aligned the number of defects it will produce is consistently close to zero. When a machine is misaligned however, the number of defects it will produce may follow a Poisson distribution (Lambert, 1992). When both the aligned and misaligned machine states are simultaneously considered, a zero-inflated dataset can result.

An additional source of zero-inflated data is within clinical trials. Measures such as the number of symptoms that a subject may exhibit or the number of risky behaviors that an individual engages in are often zero-inflated distributions (Hu et al., 2011). Finally, zero-inflated distributions can also be found in motor vehicle crash records (Lord et al., 2005). On any given day, most motor vehicles traveling on a given stretch of road will not experience a crash. The crashes that do occur follow a discrete distribution. When we consider both the motor vehicles

that do, and do not crash, we arrive at a zero-inflated distribution. The abundance of research areas that deal with zero-inflated data highlights the need for accurate models.

ANOVA

Analysis of variance (ANOVA) is a statistical method that can be used to test if population means are equal using samples from the populations in question (Fisher, 1963). Unlike the two-sample t-test, ANOVA can be used to compare the means of more than two populations. ANOVA is built on the assumptions that the data under analysis consists of independent samples, that the samples have an underlying normal distribution, and that the variance of each sample is equal. Historically, the analyses of zero-inflated data have been conducted using non-parametric rank-based tests or ANOVA. Since zero-inflation violates the assumptions of ANOVA, this can present problems. When using ANOVA in a zero-inflated context, the estimation of effects and inference can be biased (Sileshi, 2008).

Zero-Inflated Methods

Two different models for dealing with zero-inflation in linear models are zero-inflated and hurdle models. Both models rely on a binary divider between two different states of the model; the states can be thought of as ‘on’ or ‘off’. With a hurdle model, zero values can only exist in the ‘on’ state. With a zero-inflated model, zero values are possible in both the ‘on’ and ‘off’ states (Stroup, 2013). In his 2013 text, Stroup provides the following examples which nicely illustrate the difference between these two approaches:

“Zero-Inflated. In one scenario, suppose we count cars going through an intersection with a traffic light. Assuming that people obey the law, when the light

is red, cars must stop – they cannot proceed through the intersection. When we observe the system when the light is red, we always observe a zero count. When the light is green, cars are free to move through the intersection. However, even though the light is green, it is still possible that traffic is light and no cars go through the intersection. When the light is green, any count – including zero – is possible.”

“Hurdle. Suppose we count the number of times individuals visit the doctor in a given year. We have two types of individuals: those who visited the doctor during the past year and those who did not. For those who did not, the number of visits they made to the doctor must be zero. For those who did, the number of visits must be at least one. Here, the “off” state means “did not visit doctor” which necessarily means a count of zero visits. The “on” state means “did visit doctor” which necessarily means the count is one or greater. No zero counts in the “on” state.”

Additionally, Mullahy provides the following description of hurdle models:

“The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has zero or a positive realization. If the realization is positive, the ‘hurdle’ is crossed, and the conditional distribution of the positives is governed by a truncated at zero count data model” (Mullahy, 1986).

Another approach to dealing with zero-inflation is the Tobit model (Tobin, 1958) which originated for application with truncated data (Amemiya, 1973). Like the zero-inflated and hurdle models, the Tobit model utilizes an ‘on’ and ‘off’ state to differentiate between two pieces of the model. The Tobit model utilizes a latent variable which above a certain threshold is

linearly related to the dependent variable. When the latent variable is below a certain threshold, the dependent variable is always zero. While the Tobit model was originally intended for use with truncated data (McDonald & Moffitt, 1980) it can be applied to zero-inflated data (Hall & Zhang, 2004).

The 'zero-inflated Poisson' (also referred to as 'ZIP') distribution has been proposed as a solution for dealing with zero-inflation in a Poisson distribution (Greene, 1994). With reference to manufacturing machine defects, the zero-inflated Poisson distribution allows both the aligned and misaligned states to be captured by a single distribution. A parameter (p) is used to control what percentage of the time the machine is properly aligned. (Lambert, 1992). Greene also discusses the importance of distinguishing between zero-inflated and over dispersed distributions (Greene, 1994). A method for dealing with over dispersed Poisson or lack of fit to a Poisson distribution is to make likelihood adjustments (Sileshi, 2008). Though papers discussing the comparison of zero-inflated models exist (Eggers, 2015; Min & Agresti, 2002), a simulation study to quantify the relative Type I error and power while considering the variables of interest will provide additional insight.

CHAPTER 3. METHODS

In this section we will review the methods used to conduct this study. We will begin with an overview of the simulation study, followed by an explanation of the simulation parameters and how they were selected. We will then discuss our method for assessing Type I error. Finally, we will introduce the idea of ‘realized effect size’ and review our method for comparing power between the models.

Simulation Study Overview

For this simulation study, we will be estimating Type I errors for all the models when the underlying distribution means are the same. This will be done by simulating 10,000 sets of samples from the distributions and conducting a test for each set of samples. The Type I error will be estimated by counting the number of times H_0 was rejected and dividing by 10,000. We will consider three types of underlying distributions with varying levels of zero-inflation when simulating the samples. The distributions we will consider are normal, Poisson, and negative binomial. The levels of zero-inflation we will consider for each distribution are 0%, 10%, 20%, 30%, 40%, and 50%.

The number of populations we will consider are two and eight. Type I errors will be estimated first when the underlying distribution means are the same. Powers will be estimated next. Two scenarios will be considered when estimating the powers with eight populations. The first scenario is when μ_1 is different than the others ($\mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7,$ and μ_8) and the remaining 7 means are equal ($\mu_1 \neq \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$); this will be referred to as the ‘1 vs 7’ scenario. The second scenario is when the first four means ($\mu_1, \mu_2, \mu_3,$ and μ_4) are equal, and the last four means ($\mu_5, \mu_6, \mu_7,$ and μ_8) are equal but different than the first four ($\mu_1 =$

$\mu_2 = \mu_3 = \mu_4 \neq \mu_5 = \mu_6 = \mu_7 = \mu_8$); this will be referred to as the ‘4 vs 4’ scenario. In the two-population scenarios, one population is the treatment and the other is the control. In the 1 vs 7 eight-population scenario, the single population could be considered a treatment that is being compared to 7 control populations. In the 4 vs 4 eight-population scenario, 4 populations could be considered treatment populations being compared to the other 4 control populations.

Within each experimental design scenario, varying levels of sampling effort will be under consideration. For the two-population scenarios, we will consider equal sample sizes of $n = 10$, $n = 15$, $n = 20$, $n = 30$, $n = 50$, and $n = 100$ per population. For the eight-population scenarios, we will consider equal sample sizes of $n = 3$, $n = 5$, and $n = 10$ per population. For every combination of underlying distribution, level of zero-inflation, experimental design scenario, and sampling effort, 10,000 sets of samples will be generated.

During our preliminary investigation, hurdle and Tobit models were proving to be problematic for use in this study. They both exhibited poor Type I error control and power. We therefore favored additional investigation with the zero-inflated Poisson model instead, which did not exhibit these issues. Subsequently, we will consider three models for each set of samples: ANOVA, Poisson, and zero-inflated Poisson (from here on referred to as ‘ZIP’). The terms ‘ANOVA’ and ‘normal’ with reference to the model we’re considering may be used interchangeably in the remainder of this report; the ANOVA model is built on the assumption of normality. Each model will be fit and tested on its ability to discern a difference between the population means for each set of samples using SAS. PROC GENMOD will be used to fit each of the models. Minor changes to the SAS code will allow us to change the distributional assumptions between models (normal, Poisson, ZIP). Aside from changes to the distributional assumption for each model, we applied a Pearson scale to the normal and Poisson model to

address any potential issues with over dispersion. Without the Pearson over dispersion correction, Type I error (further explained below) could be affected and make power comparison of the Poisson model challenging (Stroup, 2013).

Simulation Parameters

The simulation parameters used for this study and their definitions are listed below:

Sample distribution: The underlying distribution from which samples were taken (normal, Poisson, negative binomial)

Simulations: The number of times that simulated samples were taken and models were fit to the samples

n_x : The number of data points taken during a simulation for population x .

p_x : The percentage of zero values added to the distribution for population x .

μ_x : The mean of the normal distribution (before zero-inflation is added) for population x .

σ_x : The standard deviation of the normal distribution (before zero-inflation is added) for population x .

λ_x : The lambda parameter of the Poisson distribution (before zero-inflation is added) for population x .

$k(nb)_x$: The k parameter of the negative binomial distribution (before zero-inflation is added) for population x .

$p(nb)_x$: The p parameter of the negative binomial distribution (before zero-inflation is added) for population x .

To select the underlying parameters for the three types of distributions considered in this study, we referenced the dataset that originally inspired this investigation. That dataset was from

a plant science study with a continuous response variable, which enabled us to mimic it for our simulated normal samples. We calculated a mean of 20 and a variance of 4 from the reference dataset and used those same parameters for the simulated normal samples. For consistency, we also set the lambda parameter of the Poisson simulated samples to 20. This value of lambda kept the shape of the Poisson samples symmetric and relatively normal. Likewise, with the negative binomial parameters, we set $k(nb)$ to 20 and $p(nb)$ to 0.5. These parameters kept the shape of the negative binomial samples symmetric and relatively normal. Setting the Poisson and negative binomial parameters so that the distributions are symmetric and relatively normal should reduce any potential effects of distribution skewness on our results. Though the variance is not the same between the three distributions, the mean is 20 for each. Histograms of each distribution with zero-inflation of $p = 0.1$ (10% zero-inflation) can be found in Figures 1, 2, and 3 below. Note that although a zero-inflation simulation parameter of $p = 0.1$ (10% zero-inflation) was used for each distribution shown below, the actual level of zero-inflation present in the samples can vary based on random chance.

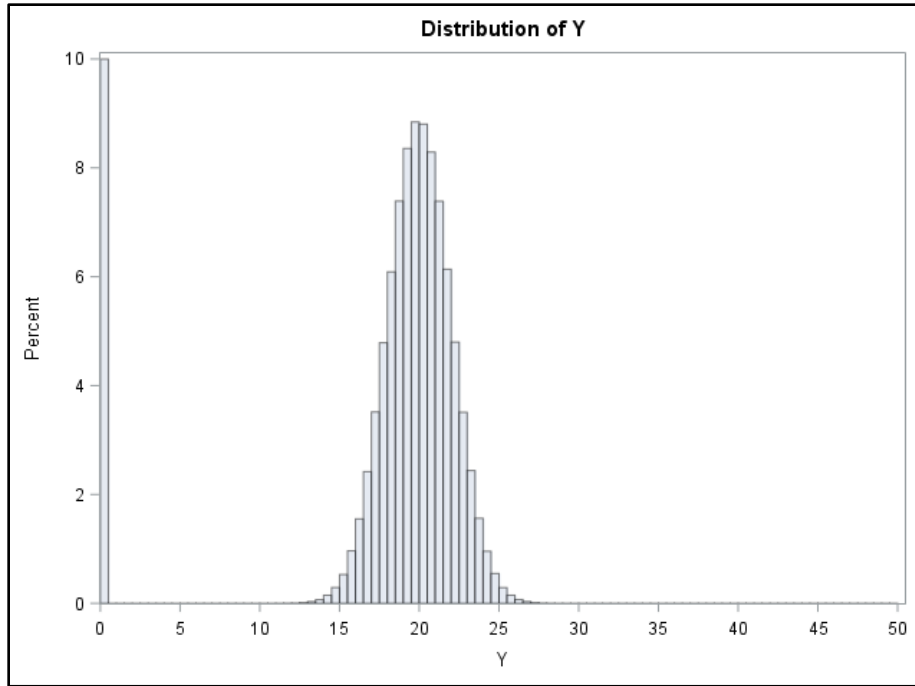


Figure 1. Normal Distribution – $\mu = 20$, $\sigma = 2$, $p = 0.1$.

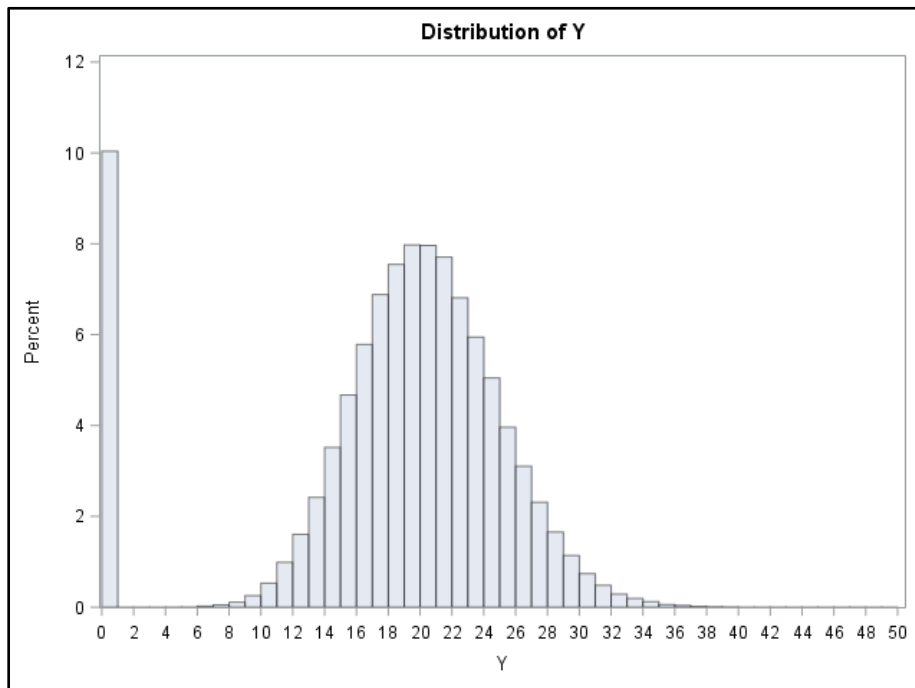


Figure 2. Poisson Distribution – $\lambda = 20$, $p = 0.1$.

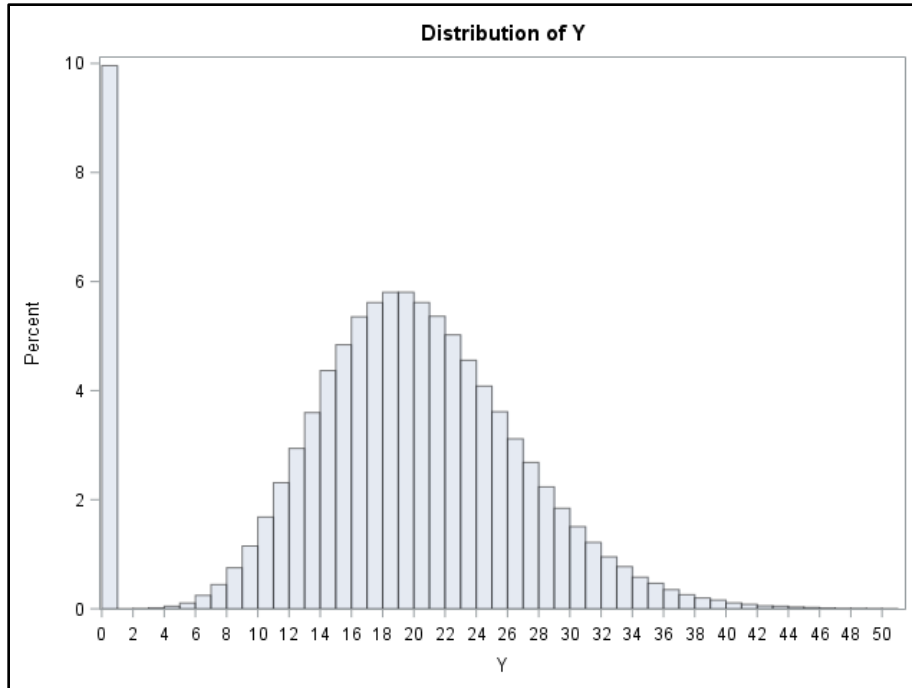


Figure 3. Negative Binomial Distribution – $k(nb) = 20$, $p(nb) = 0.5$, $p = 0.1$.

Type I Error Assessment

When conducting a comparative power analysis, it is important to verify that each model being compared has controlled Type I error. Comparing the power between two models that do not have similar Type I error rates could lead to incorrect conclusions about which model is more powerful; seemingly small deviations in Type I error can cause significant changes in apparent power.

For this study, we estimated Type I error for the two-population scenarios by simulating two samples from the same distribution, fitting the models to each pair of samples, and tallying rejections to the null and alternative hypothesis below:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

The Type I error for each model was estimated by taking the number of samples that resulted in a rejection of the null hypothesis and dividing by 10,000 (the number of sample pairs simulated).

We used a stated alpha value (α) of 0.05 for this study. Type I error was estimated for each combination of sample distribution (normal, Poisson, negative binomial), model, sampling effort, and level of zero-inflation.

For the eight-population scenarios, we compared Type I error by conducting a test with different null and alternative hypothesis from the two-population scenarios. The null and alternative hypothesis for the eight-population treatment scenarios are as follows:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$$

$$H_a: \text{At least one } \mu_x \text{ is different from the others}$$

As with the two-population case, Type I error for each model was estimated by taking the number of samples that resulted in a rejection of the null hypothesis and dividing by 10,000 (the number of sample pairs simulated). Type I error was estimated for each combination of sample distribution (normal, Poisson, negative binomial), model, sampling effort, and level of zero-inflation.

Realized Effect Size

We next conducted a power analysis. To do this, it is important to understand the impact that zero-inflation has on treatment effect size. Consider a sample from a normal distribution with mean 20 and variance 4. Without any zero-inflation, the expected value for the sample mean and variance are respectively 20 and 4. Now let's consider another sample in which half of the values are from the same normal distribution, and the other half of the values are zeros (50% zero-inflation, $p = 0.5$). For this sample, the expected value for the mean and variance are respectively 10 and 102.

To understand this within the context of effect size, let's consider a normally distributed population with mean 22 and variance 4, and another population with mean 20 and variance 4. The normalized effect size between these two populations is 1σ $((22 - 20) / 2)$. If we apply 50% zero-inflation to those samples, the means and variance will change significantly, which will also change the normalized effect size between the two populations. For the purposes of this study, we have defined the term 'realized effect size' as the difference between the zero-inflated means divided by the variance before zero-inflation was added. While this may be a somewhat overly simplistic approach, it succinctly illustrates the point of decreasing effect size on the non-zero portion of the data as zero-inflation increases. Applying this calculation to the previously mentioned normal populations, with means 22 and 20 respectively, we are left with a realized effect size of 0.5σ $((11 - 10) / 2 = 0.5\sigma)$. In Table 1 below, the realized effect size for each level of zero-inflation and target effect size has been calculated. It is important to keep in mind the diminishing realized effect size as zero-inflation increases because we expect lower statistical power on samples from populations with a smaller effect size.

Table 1. Realized Effect Size.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|-------------------------|---------|---------|---------|---------|---------|---------|
| 1σ Realized Effect Size | 1.0σ | 0.9σ | 0.8σ | 0.7σ | 0.6σ | 0.5σ |
| 2σ Realized Effect Size | 2.0σ | 1.8σ | 1.6σ | 1.4σ | 1.2σ | 1.0σ |

Power Comparison

The purpose of a power comparison is to determine if certain models perform better under specific circumstances. In general, a statistical test will have higher power when the effect size is large. To quantify the difference between means, the term 'standardized effect size' is used and refers to the difference in population means divided by the standard deviation of the

samples. For example, if we were sampling from the distributions $N(22,4)$ and $N(20,4)$, where the variance of each distribution is equal to 4 (standard deviation equal to 2), we would have an effect size of $(22 - 20) / 2 = 1\sigma$. For the purposes of this study, we considered effect sizes of 1σ and 2σ for each of the experimental design scenarios.

For the two-population scenarios in this study, we compared power by simulating two samples from distributions with a set effect size, fitting the models to the samples, and tallying rejections to the null and alternative hypothesis below:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Since the two samples were taken from distributions with different means, the models arrive at the correct conclusion when they reject H_0 and conclude that there is a difference between the population means. For each combination of sample distribution (normal, Poisson, negative binomial), model, sampling effort, and level of zero-inflation, we tallied results for 10,000 samples of the test described above. An overall power for each scenario was calculated by taking the number of samples that resulted in a rejection of the null hypothesis and dividing by 10,000.

For the eight-population scenarios, we compared power by conducting a test with different null and alternative hypothesis from the two-population scenarios. The null and alternative hypothesis for both eight-population treatment scenarios are as follows:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$$

$$H_a: \text{At least one } \mu_x \text{ is different from the others}$$

Like the two-population scenarios, the eight-population scenarios arrived at the correct conclusion when H_0 was rejected to conclude that at least one of the groups had a different population mean.

Again, for each combination of sample distribution (normal, Poisson, negative binomial), model, sampling effort, and level of zero-inflation, we tallied results for 10,000 samples of the test described above. An overall power for each scenario was calculated by taking the number of samples that resulted in a rejection of the null hypothesis and dividing by 10,000.

Simulation Study Outline

To more succinctly describe this simulation study, we have created the outline below:

- 1) Type I error assessment with two populations, $\mu_1 = \mu_2$.
 - a) Underlying distributions: Normal, Poisson, Negative Binomial
 - b) Zero-inflation of 0%, 10%, 20%, 30%, 40%, and 50%
 - c) Sampling effort of $n = 10$, $n = 15$, $n = 20$, $n = 30$, $n = 50$, $n = 100$ (per population)
 - d) Models: ANOVA, Poisson, Zero-Inflated Poisson
 - e) Samples taken / models fit for each combination of parameters above: 10,000
- 2) Type I error assessment with eight populations, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$.
 - a) Underlying distributions: Normal, Poisson, Negative Binomial
 - b) Zero-inflation of 0%, 10%, 20%, 30%, 40%, and 50%
 - c) Sampling effort of $n = 3$, $n = 5$, $n = 10$ (per population)
 - d) Models: ANOVA, Poisson, Zero-Inflated Poisson
 - e) Samples taken / models fit for each combination of parameters above: 10,000
- 3) Power comparison with two populations, $\mu_1 \neq \mu_2$.
 - a) Underlying distributions: Normal, Poisson, Negative Binomial
 - b) Zero-inflation of 0%, 10%, 20%, 30%, 40%, and 50%
 - c) Sampling effort of $n = 10$, $n = 15$, $n = 20$, $n = 30$, $n = 50$, $n = 100$ (per population)
 - d) Models: ANOVA, Poisson, Zero-Inflated Poisson

- e) Effect sizes of 1σ and 2σ
 - f) Samples taken / models fit for each combination of parameters above: 10,000
- 4) Power comparison with eight populations, $\mu_1 \neq \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$.
- a) Underlying distributions: Normal, Poisson, Negative Binomial
 - b) Zero-inflation of 0%, 10%, 20%, 30%, 40%, and 50%
 - c) Sampling effort of $n = 3$, $n = 5$, $n = 10$ (per population)
 - d) Models: ANOVA, Poisson, Zero-Inflated Poisson
 - e) Effect sizes of 1σ and 2σ
 - f) Samples taken / models fit for each combination of parameters above: 10,000
- 5) Power comparison with eight populations, $\mu_1 = \mu_2 = \mu_3 = \mu_4 \neq \mu_5 = \mu_6 = \mu_7 = \mu_8$.
- a) Underlying distributions: Normal, Poisson, Negative Binomial
 - b) Zero-inflation of 0%, 10%, 20%, 30%, 40%, and 50%
 - c) Sampling effort of $n = 3$, $n = 5$, $n = 10$ (per population)
 - d) Effect sizes of 1σ and 2σ
 - e) Samples taken / models fit for each combination of parameters above: 10,000

SAS Code

To reference the SAS code used for this study, see Appendix C. For the two-population scenarios, portions of the eight-population code in Appendix C were commented out.

CHAPTER 4. RESULTS

In this section we will provide an overview of the results from the Type I error assessment and review the results from the power comparison for each sample distribution we considered.

Two-Population Type I Error Assessment

As stated in the ‘Methods’ section, because a target Type I error rate of $\alpha = 0.05$ was selected for this study, values for Type I error should be near 0.05. To create a reference for the interpretation of our Type I error rates, we calculated a 95% confidence interval on the binomial proportion with the point estimator set to 0.05 (and $n = 10,000$); the resulting interval is [0.0457, 0.0543]. For comparison, the 90% and 99% intervals are respectively [0.0464, 0.0536] and [0.0444, 0.0556].

The results of the two-population Type I error assessment on normal samples can be found in Tables 2, 3, and 4 below. In these tables, we see Type I error maintained near the stated level of alpha for the normal model. The Poisson model also has Type I error maintained near the stated level of alpha for samples of $n = 15$ and larger. This model is exhibiting conservative Type I error for samples of $n = 10$ (except for the $p = 0.0$ case). The ZIP model applied to zero-inflated normal data has a consistent Type I error rate near zero for each of the test cases.

Table 2. Normal Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.95 | 2.82 | 4.78 | 4.92 | 4.98 | 5.06 |
| n = 15 | 4.94 | 4.09 | 4.63 | 5.55 | 4.99 | 5.29 |
| n = 20 | 4.95 | 4.52 | 4.94 | 4.81 | 5.10 | 4.87 |
| n = 30 | 5.03 | 4.36 | 4.70 | 5.30 | 4.71 | 5.10 |
| n = 50 | 4.94 | 4.65 | 5.11 | 4.93 | 4.88 | 5.23 |
| n = 100 | 5.05 | 5.02 | 4.95 | 4.74 | 5.14 | 4.90 |

Table 3. Normal Sample – Poisson Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.54 | 2.43 | 3.51 | 3.84 | 3.72 | 3.29 |
| n = 15 | 5.64 | 3.64 | 4.53 | 4.45 | 3.80 | 3.76 |
| n = 20 | 6.03 | 4.10 | 4.58 | 4.70 | 4.70 | 4.20 |
| n = 30 | 5.50 | 4.34 | 4.47 | 4.77 | 4.49 | 4.59 |
| n = 50 | 5.47 | 4.23 | 4.77 | 4.43 | 5.09 | 4.87 |
| n = 100 | 5.21 | 5.00 | 4.65 | 4.83 | 5.04 | 5.42 |

Table 4. Normal Sample – ZIP Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| n = 20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| n = 30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 100 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |

The results of the two-population Type I error analysis on zero-inflated Poisson samples can be found in Tables 5, 6, and 7 below. For the normal model, Type I error is being controlled across sampling effort and zero-inflation.

Table 5. Poisson Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 5.07 | 4.19 | 4.85 | 4.97 | 5.02 | 4.97 |
| n = 15 | 5.02 | 4.86 | 5.42 | 4.81 | 5.14 | 4.88 |
| n = 20 | 4.97 | 4.82 | 4.88 | 5.16 | 4.77 | 5.13 |
| n = 30 | 5.07 | 5.14 | 5.21 | 5.26 | 5.20 | 5.01 |
| n = 50 | 4.57 | 5.22 | 5.12 | 5.19 | 5.06 | 5.02 |
| n = 100 | 5.10 | 5.40 | 4.84 | 5.10 | 4.97 | 4.96 |

Table 6. Poisson Sample – Poisson Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.42 | 4.80 | 4.45 | 4.15 | 4.10 | 3.53 |
| n = 15 | 5.96 | 5.10 | 4.96 | 4.15 | 4.02 | 4.16 |
| n = 20 | 5.73 | 4.61 | 4.78 | 4.87 | 4.62 | 4.49 |
| n = 30 | 5.54 | 5.08 | 5.25 | 4.75 | 4.84 | 4.22 |
| n = 50 | 5.02 | 4.74 | 4.93 | 4.80 | 4.91 | 5.01 |
| n = 100 | 5.31 | 5.18 | 5.00 | 4.81 | 4.42 | 4.76 |

Table 7. Poisson Sample – ZIP Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.86 | 4.73 | 4.95 | 4.85 | 4.92 | 4.57 |
| n = 15 | 4.78 | 5.07 | 4.87 | 5.11 | 4.71 | 4.85 |
| n = 20 | 5.00 | 4.67 | 4.89 | 5.11 | 4.78 | 5.26 |
| n = 30 | 5.04 | 4.89 | 4.95 | 4.68 | 4.64 | 5.00 |
| n = 50 | 4.77 | 4.82 | 4.95 | 4.78 | 5.04 | 4.92 |
| n = 100 | 5.15 | 5.35 | 4.89 | 5.08 | 4.87 | 4.71 |

The results of the two-population Type I error analysis on zero-inflated negative binomial samples can be found in Tables 8, 9, and 10 below. The normal and Poisson models have expected Type I error across all levels of zero-inflation and sampling effort. With the ZIP model,

Type I error is unacceptably high and subsequently this model will not be suitable for power comparison against the normal and Poisson models.

Table 8. Negative Binomial Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.80 | 5.48 | 4.80 | 5.58 | 4.82 | 4.88 |
| n = 15 | 5.15 | 4.66 | 4.85 | 4.64 | 4.74 | 5.13 |
| n = 20 | 4.74 | 5.13 | 4.82 | 5.26 | 5.17 | 4.99 |
| n = 30 | 4.84 | 5.27 | 4.64 | 5.25 | 4.84 | 5.18 |
| n = 50 | 5.13 | 4.82 | 4.86 | 5.16 | 4.99 | 4.94 |
| n = 100 | 5.06 | 4.98 | 4.95 | 4.97 | 4.70 | 5.09 |

Table 9. Negative Binomial Sample – Poisson Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.65 | 5.76 | 5.06 | 4.09 | 4.51 | 3.74 |
| n = 15 | 5.61 | 5.19 | 4.81 | 4.78 | 4.45 | 4.38 |
| n = 20 | 5.66 | 5.31 | 4.87 | 5.25 | 4.54 | 4.81 |
| n = 30 | 5.03 | 5.20 | 5.02 | 4.82 | 4.39 | 4.99 |
| n = 50 | 5.46 | 4.80 | 4.91 | 4.73 | 4.62 | 4.69 |
| n = 100 | 5.42 | 4.76 | 5.03 | 5.18 | 4.92 | 5.00 |

Table 10. Negative Binomial Sample – ZIP Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 16.75 | 16.42 | 17.18 | 15.81 | 15.69 | 16.43 |
| n = 15 | 16.62 | 16.11 | 16.38 | 16.12 | 16.01 | 16.01 |
| n = 20 | 17.19 | 16.56 | 16.46 | 16.31 | 17.02 | 16.04 |
| n = 30 | 15.69 | 16.21 | 15.99 | 16.05 | 16.35 | 17.23 |
| n = 50 | 17.01 | 16.82 | 16.45 | 16.03 | 17.17 | 16.36 |
| n = 100 | 16.84 | 16.53 | 16.25 | 16.20 | 16.23 | 16.45 |

Eight-Population Type I Error Assessment

The results of the eight-population Type I error assessment on normal samples can be found in Tables 11, 12, and 13 below. For the normal model, we see maintained Type I error across levels of sampling size and zero-inflation. The $n = 10$ case for the Poisson model applied to normal data is borderline acceptable and power results will need to be interpreted with caution. In Table 13 we see near zero Type I error for all levels of sample size and zero-inflation for the ZIP model applied to normal samples.

Table 11. Normal Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 4.89 | 4.33 | 5.12 | 5.21 | 6.01 | 5.62 |
| $n = 5$ | 5.17 | 4.20 | 5.34 | 4.97 | 5.26 | 5.34 |
| $n = 10$ | 4.93 | 4.16 | 5.13 | 4.82 | 5.32 | 4.86 |

Table 12. Normal Sample – Poisson Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 12.14 | 2.51 | 5.93 | 12.18 | 14.68 | 10.79 |
| $n = 5$ | 8.62 | 2.64 | 2.86 | 3.99 | 7.59 | 11.42 |
| $n = 10$ | 6.33 | 3.58 | 3.78 | 4.06 | 4.39 | 4.23 |

Table 13. Normal Sample – ZIP Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 0.00 | 0.17 | 0.40 | 0.63 | 1.24 | 1.28 |
| $n = 5$ | 0.00 | 0.00 | 0.13 | 0.41 | 0.33 | 0.78 |
| $n = 10$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.31 |

The results of the eight-population Type I error analysis on zero-inflated Poisson samples can be found in Tables 14, 15 and 16 below. In Table 15 we see inflated or deflated Type I error

except for the $n = 10$ case with the Poisson model. In Table 16 we see Type I error only being controlled for the $n = 10$ cases for the ZIP model.

Table 14. Poisson Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 5.22 | 4.71 | 5.11 | 5.74 | 5.30 | 5.19 |
| $n = 5$ | 4.83 | 4.51 | 4.75 | 4.32 | 5.43 | 5.15 |
| $n = 10$ | 4.86 | 4.93 | 5.32 | 5.19 | 4.89 | 5.55 |

Table 15. Poisson Sample – Poisson Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 12.02 | 4.57 | 6.96 | 12.32 | 14.18 | 12.24 |
| $n = 5$ | 8.38 | 3.98 | 4.33 | 4.54 | 7.39 | 10.70 |
| $n = 10$ | 6.35 | 4.51 | 4.22 | 4.50 | 4.86 | 5.11 |

Table 16. Poisson Sample – ZIP Model Eight-Population Type I Error.

| Zero-Inflation | $p = 0.0$ | $p = 0.1$ | $p = 0.2$ | $p = 0.3$ | $p = 0.4$ | $p = 0.5$ |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $n = 3$ | 5.42 | 5.78 | 8.08 | 11.21 | 13.92 | 16.99 |
| $n = 5$ | 5.22 | 4.79 | 5.22 | 6.21 | 8.34 | 11.78 |
| $n = 10$ | 4.47 | 5.36 | 5.10 | 4.97 | 5.23 | 5.88 |

The results of the eight-population Type I error analysis on zero-inflated negative binomial samples can be found in Tables 17, 18, and 19 below. In Table 18 we again see unacceptable Type I error for the Poisson models except for the $n = 10$ cases. In Table 19 we see very high Type I error for the ZIP model. In this case, Type I error is unacceptably high and subsequently this model will not be suitable for power comparison against the normal and Poisson models. All Type I error assessment results and the simulation parameters used to calculate them are available in Appendix A.

Table 17. Negative Binomial Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 4.75 | 4.82 | 5.28 | 5.26 | 5.46 | 5.33 |
| n = 5 | 4.88 | 4.99 | 5.16 | 4.81 | 4.99 | 5.32 |
| n = 10 | 4.68 | 4.55 | 5.67 | 5.18 | 5.27 | 5.30 |

Table 18. Negative Binomial Sample – Poisson Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.41 | 7.01 | 8.69 | 12.82 | 14.67 | 13.20 |
| n = 5 | 8.31 | 5.76 | 4.87 | 5.81 | 8.61 | 11.04 |
| n = 10 | 6.64 | 4.96 | 5.34 | 5.05 | 4.95 | 5.65 |

Table 19. Negative Binomial Sample – ZIP Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 42.61 | 42.88 | 45.24 | 49.41 | 52.41 | 55.01 |
| n = 5 | 42.63 | 42.31 | 41.73 | 43.32 | 45.90 | 48.49 |
| n = 10 | 42.50 | 43.12 | 43.43 | 42.35 | 43.38 | 42.98 |

Zero-Inflated Normal Power Comparison

For two-population zero-inflated normal samples, we see the ZIP model having stronger power than the normal or Poisson models for sampling efforts above $n = 30$. Without zero inflation ($p = 0.0$) the normal and Poisson models have higher power than the ZIP model. It is worth noting that the ZIP model applied to two-population normal samples had low Type I error which suggests that the power shown in the figures below could improve if the Type I error were closer to the desired 5%. The 1σ effect size power comparison plots for the normal, Poisson, and ZIP models applied to zero-inflated normal samples can be found in Figures 4, 5, 6, 7, 8, and 9 below. The relative power of the models at the 1σ effect size were amplified in the 2σ effect size simulations. To review power plots for the 2σ effect size, see Appendix B.

For eight-population zero-inflated normal samples, we only saw acceptable Type I error for samples of $n = 10$. Subsequently, we will only compare models for this distribution at the $n = 10$ sample size. Reviewing Figures 10, 11, 12, 13, 14, and 15 below, we see that there is no distinct power advantage for any of the models. This power result did not vary when the treatment was applied to 1 population (1 vs 7) versus when the treatment was applied to 4 populations (4 vs 4). To review power plots for all combinations of effect size and treatment group scenario, see Appendix B.

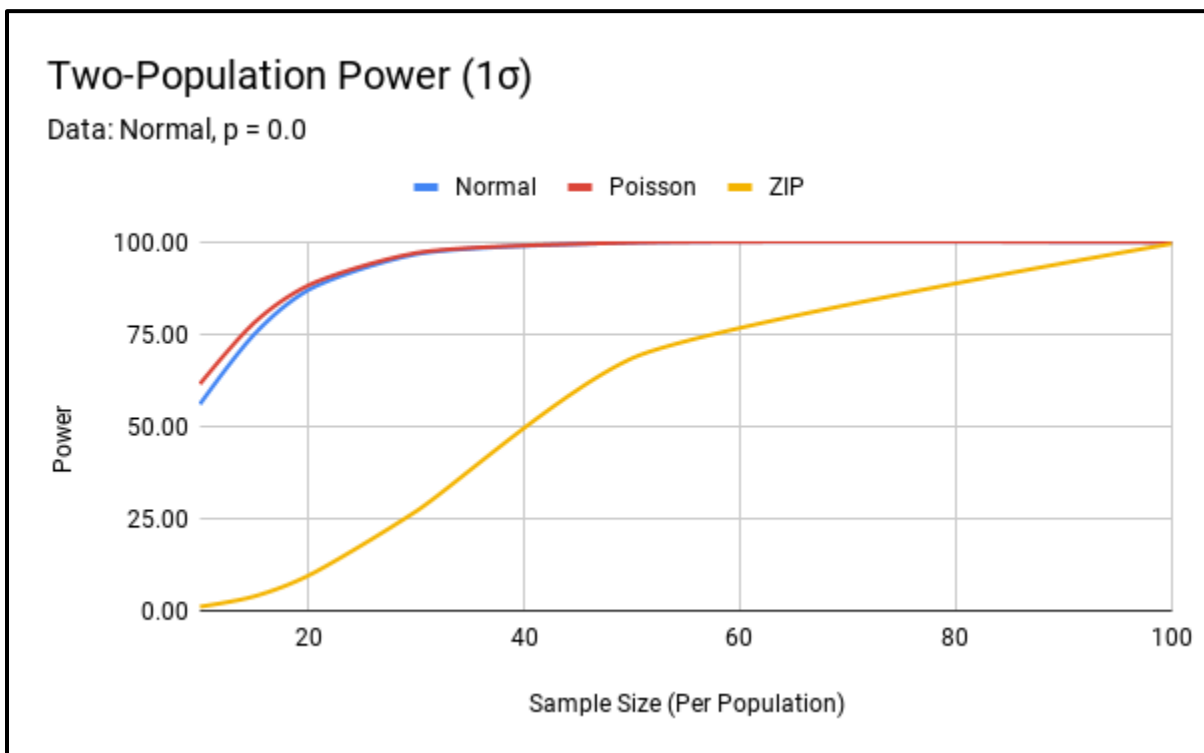


Figure 4. Normal Sample – $p = 0.0$ Two-Population Power (1σ).

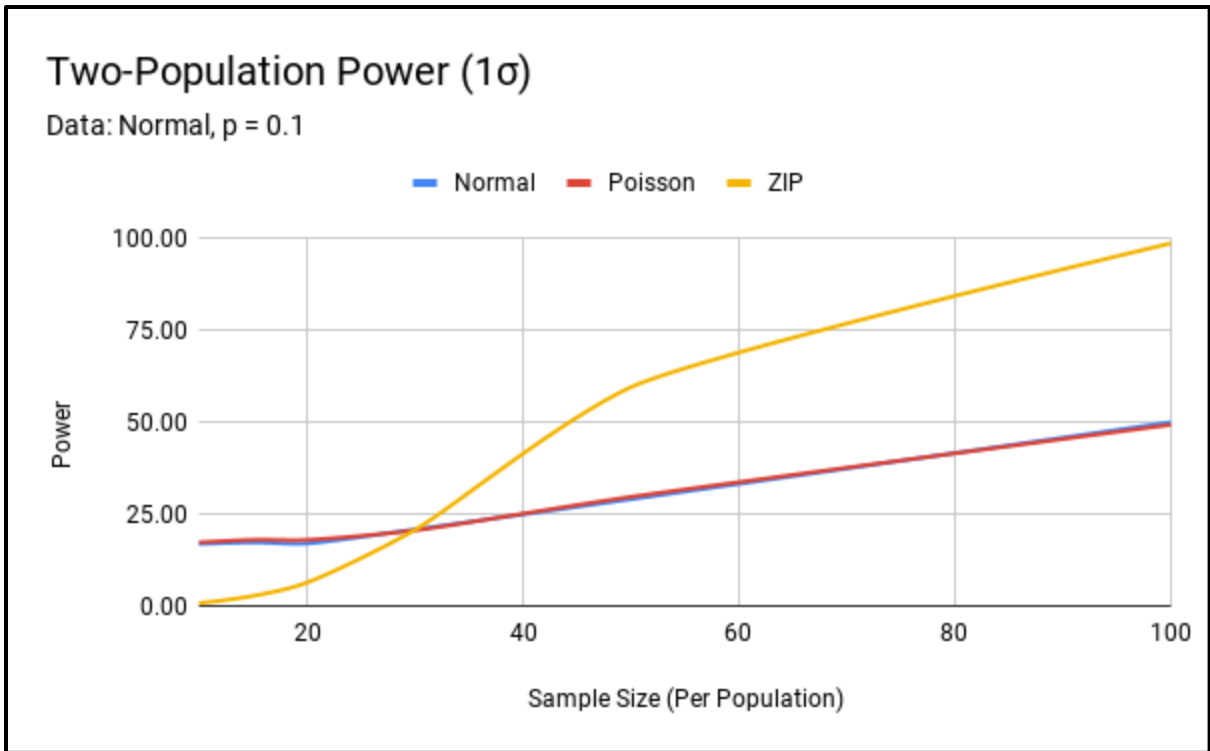


Figure 5. Normal Sample – $p = 0.1$ Two-Population Power (1σ).

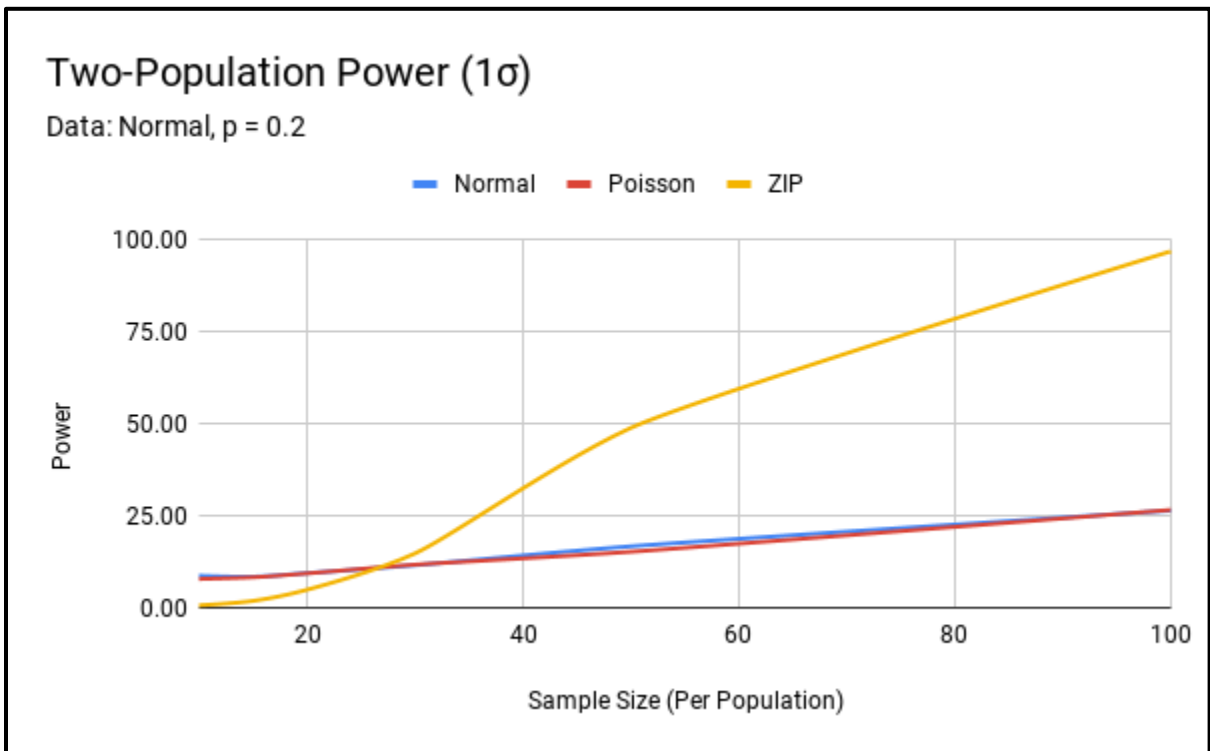


Figure 6. Normal Sample – $p = 0.2$ Two-Population Power (1σ).

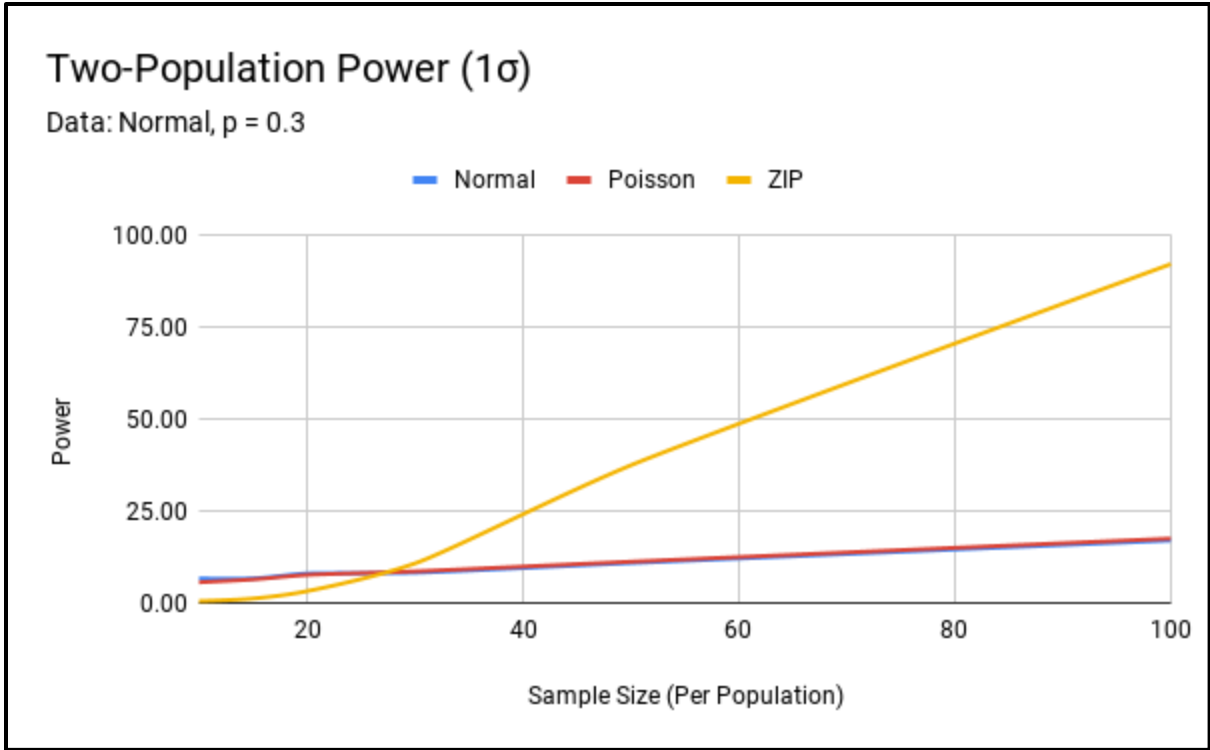


Figure 7. Normal Sample – $p = 0.3$ Two-Population Power (1σ).

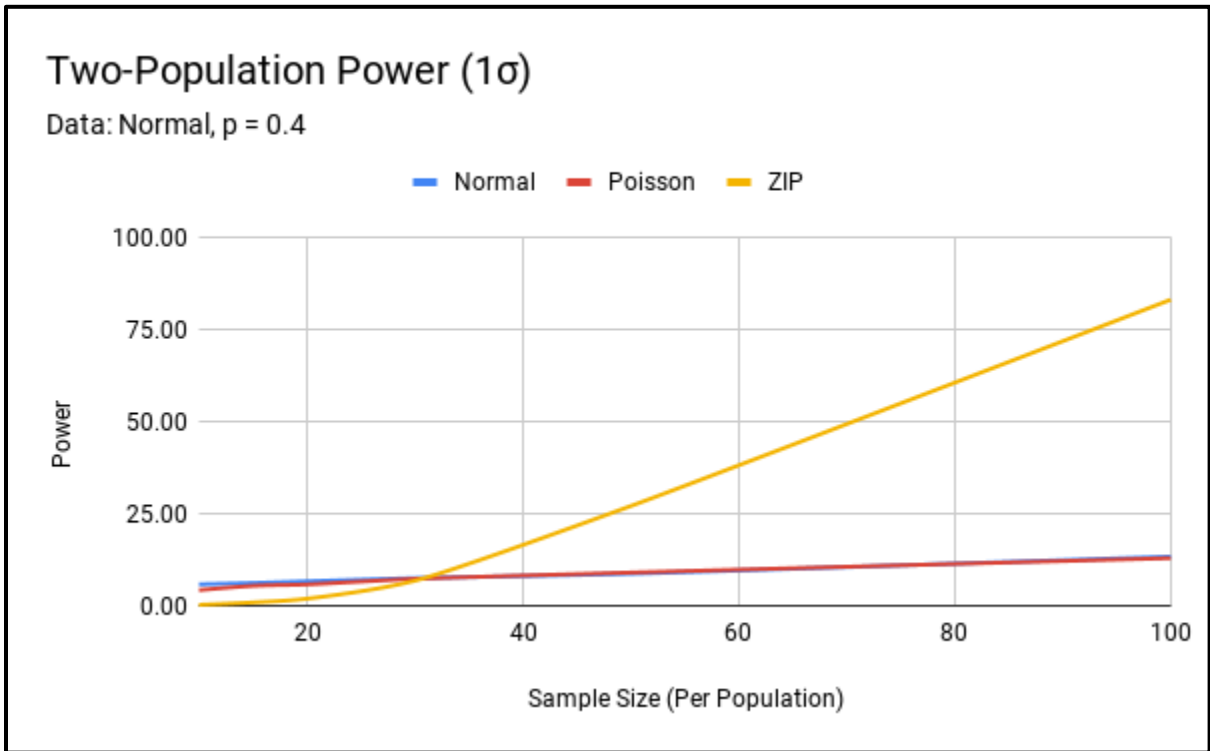


Figure 8. Normal Sample – $p = 0.4$ Two-Population Power (1σ).

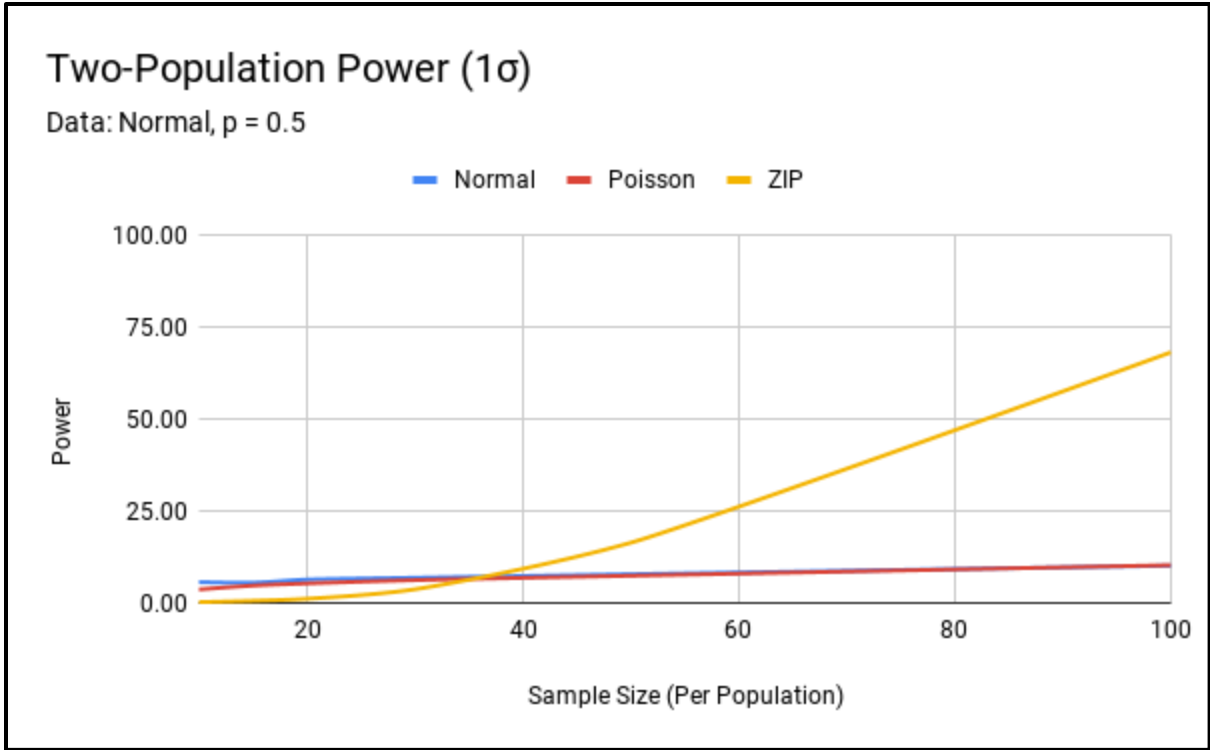


Figure 9. Normal Sample – $p = 0.5$ Two-Population Power (1σ).

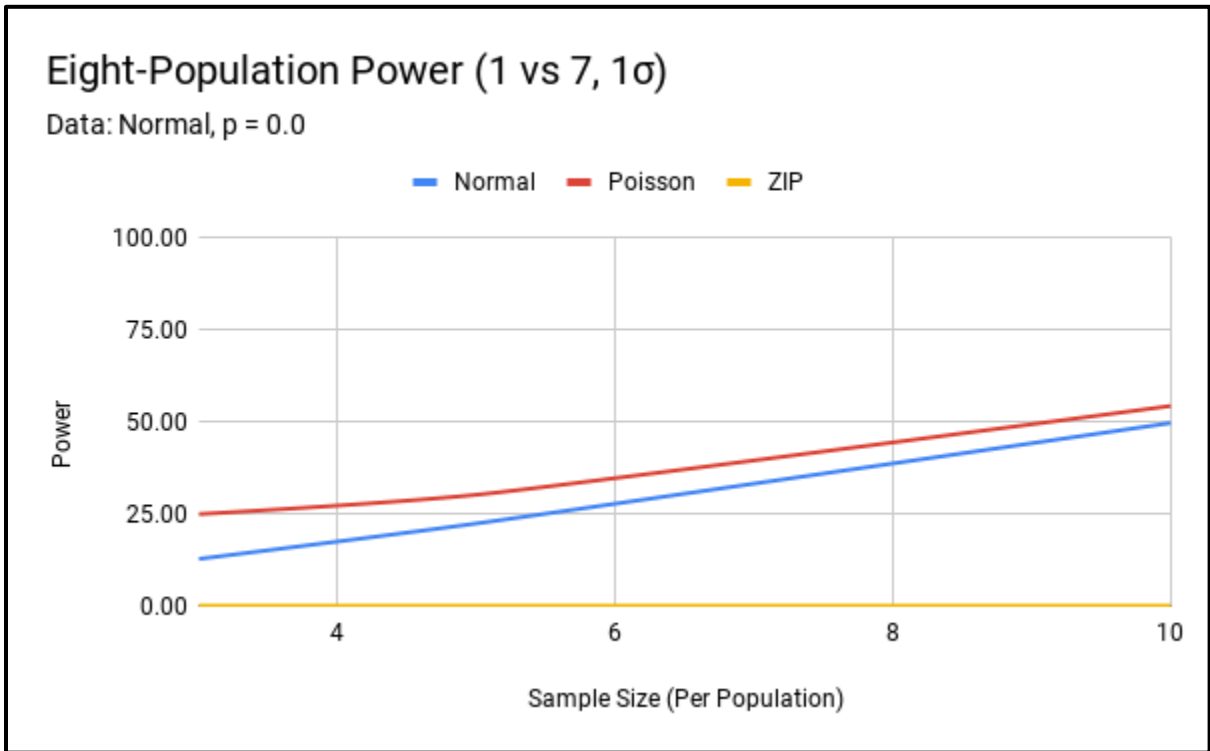


Figure 10. Normal Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ).

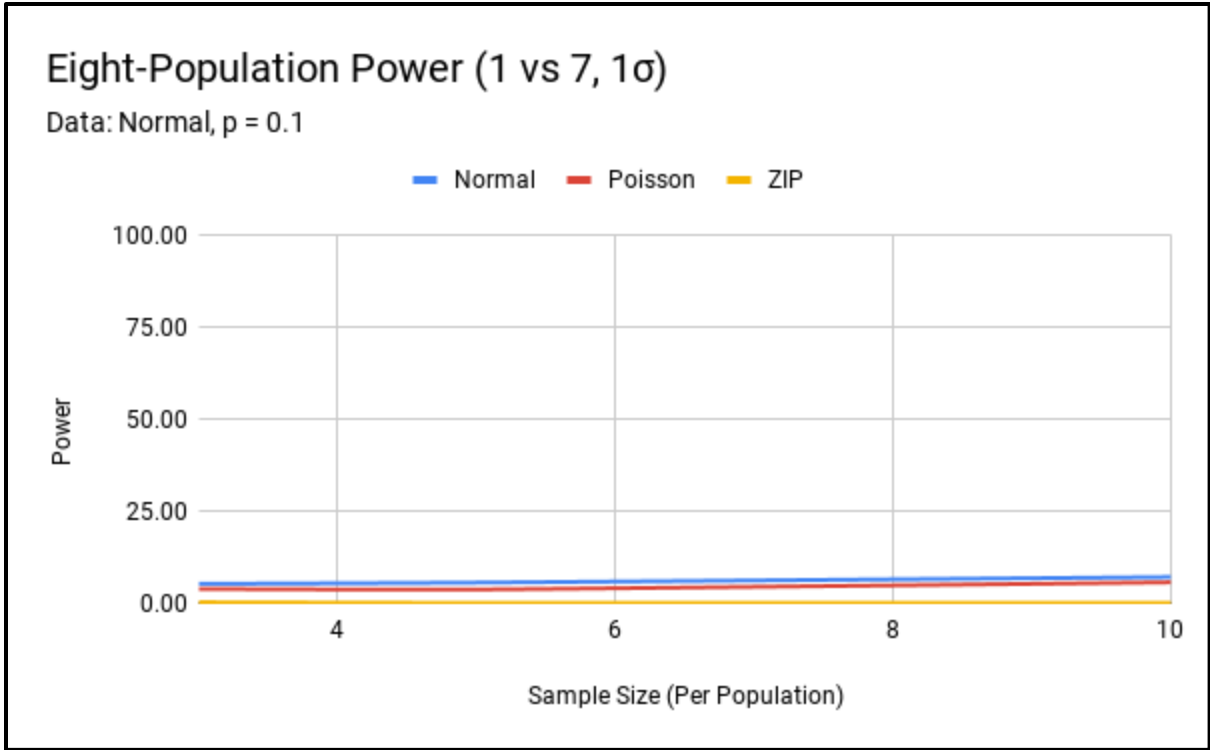


Figure 11. Normal Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1 σ).

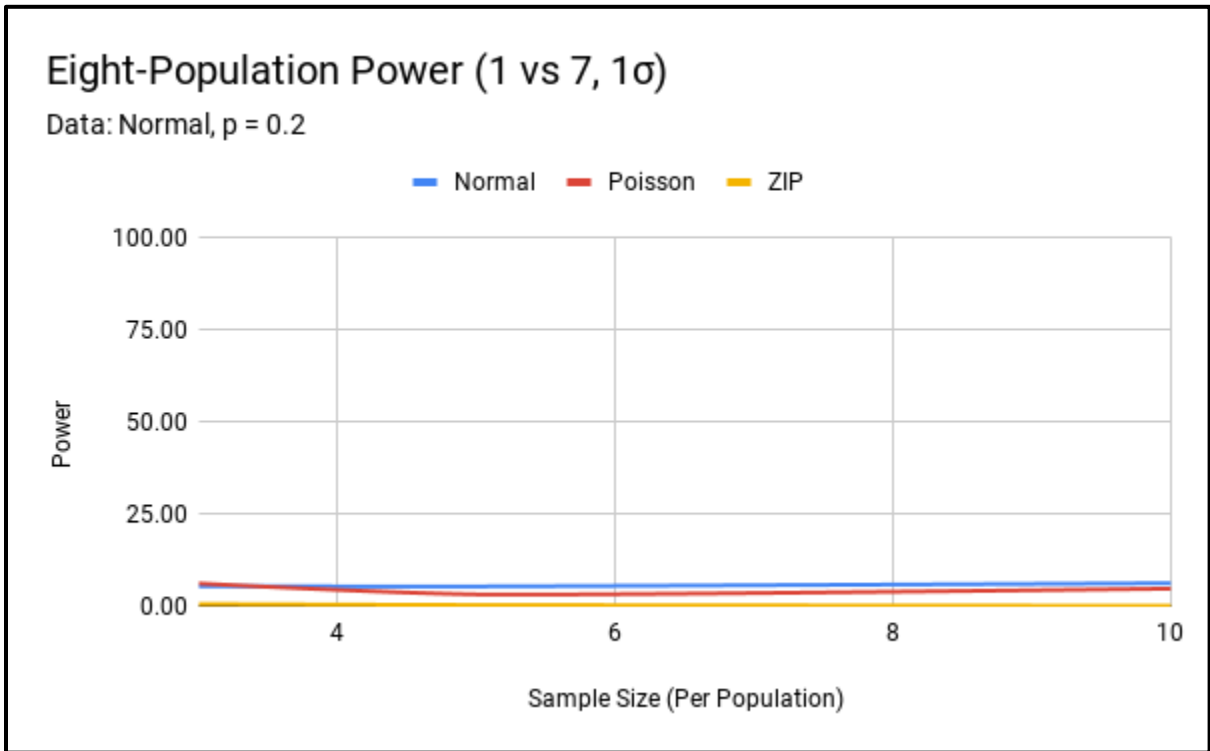


Figure 12. Normal Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1 σ).

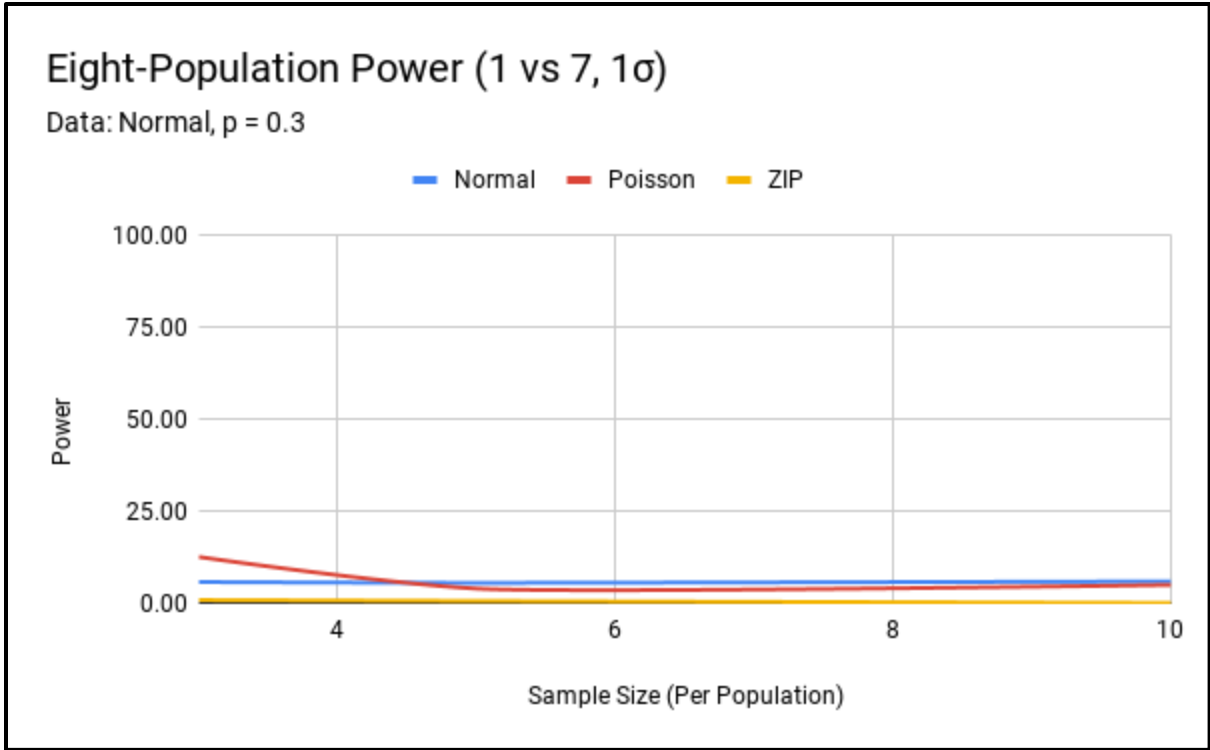


Figure 13. Normal Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1σ).

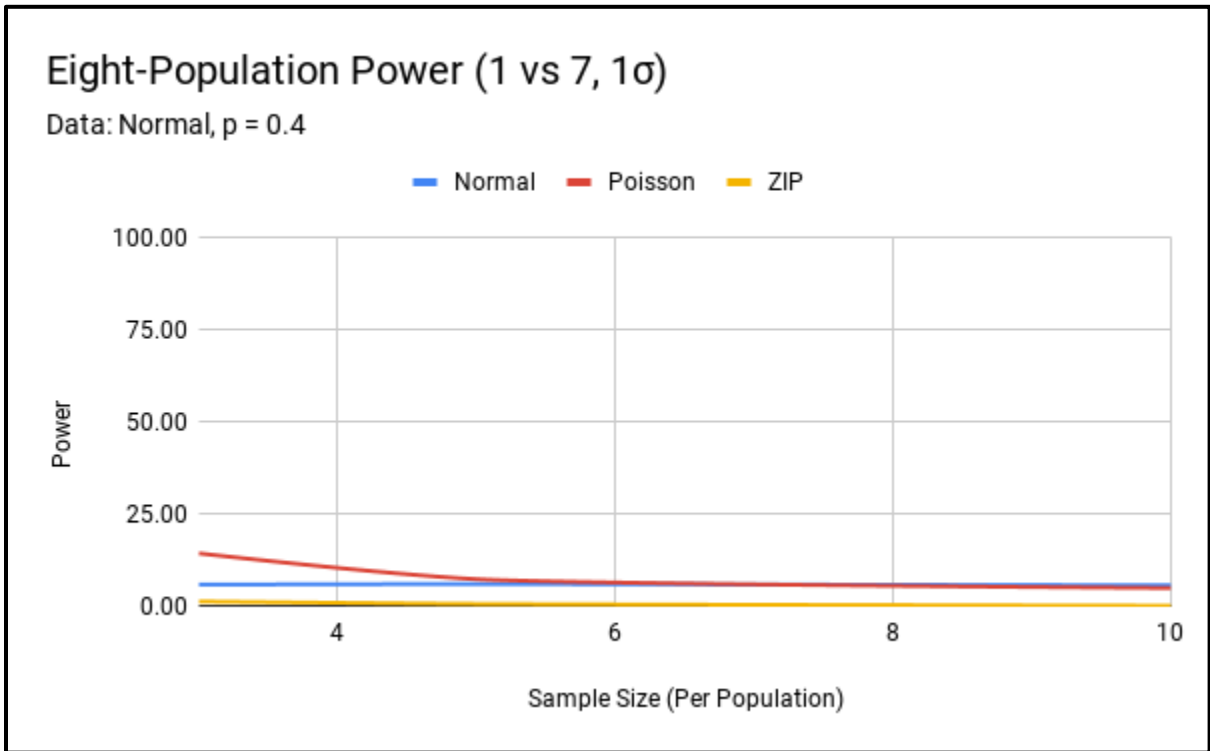


Figure 14. Normal Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1σ).

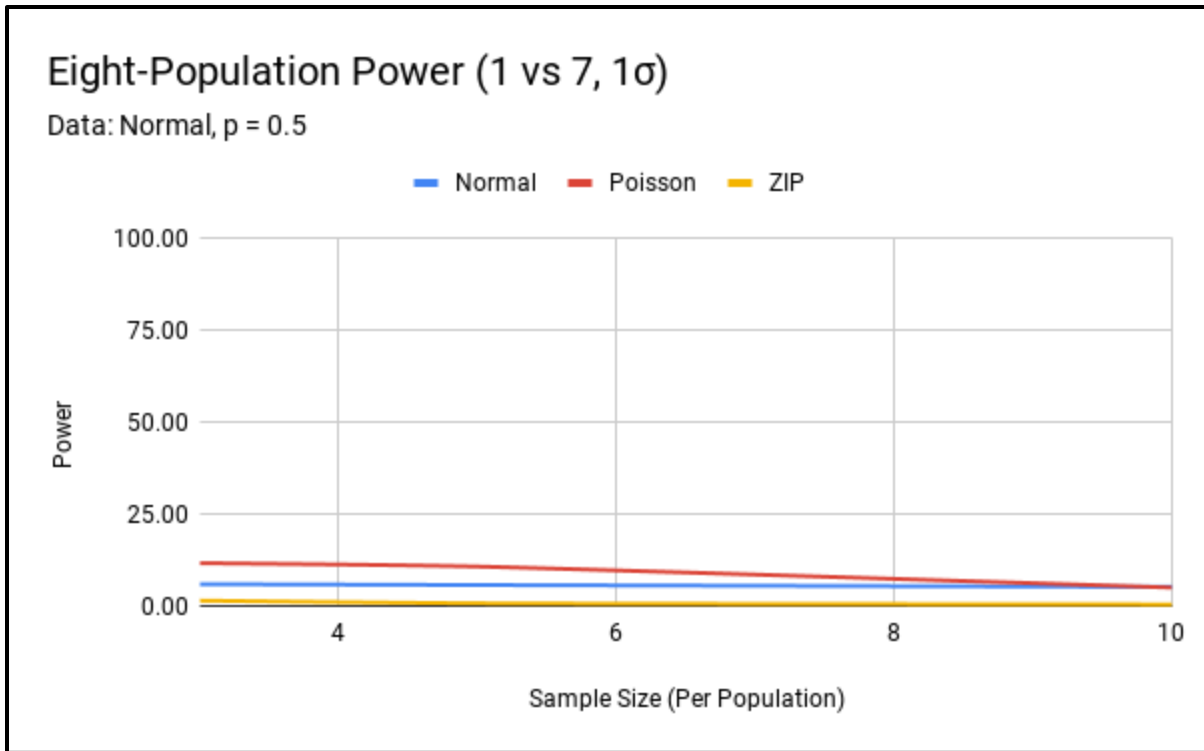


Figure 15. Normal Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ).

Zero-Inflated Poisson Power Comparison

For two-population zero-inflated Poisson samples, the normal and Poisson models had comparable power across levels of zero-inflation and sampling effort. Both the normal and Poisson models were however outperformed by the ZIP model for all combinations of zero-inflation and sampling effort. The 1σ effect size power comparison plots for the normal, Poisson, and ZIP models applied to zero-inflated Poisson samples can be found in Figures 16, 17, 18, 19, 20, and 21 below. The relative power of the models at the 1σ effect size were amplified in the 2σ effect size simulations. To review power plots for the 2σ effect size, see Appendix B.

For eight-population zero-inflated Poisson samples, we only saw acceptable Type I error for samples of $n = 10$. Subsequently, we will only compare models for this distribution at the $n = 10$ sample size. We again saw there was power improvement with the ZIP model over the normal

and Poisson models. The normal and Poisson models again performed similarly. The relative strength of the models did not change for the 1 population (1 vs 7) or 4 population (4 vs 4) treatment group scenarios, but there was improved power for all models with the 4 population treatment group scenarios. Power comparison plots for the 1 population treatment samples (1 vs 7) can be found in Figures 22, 23, 24, 25, 26, and 27 below. To review power plots for all combinations of effect size and treatment group scenario, see Appendix B.

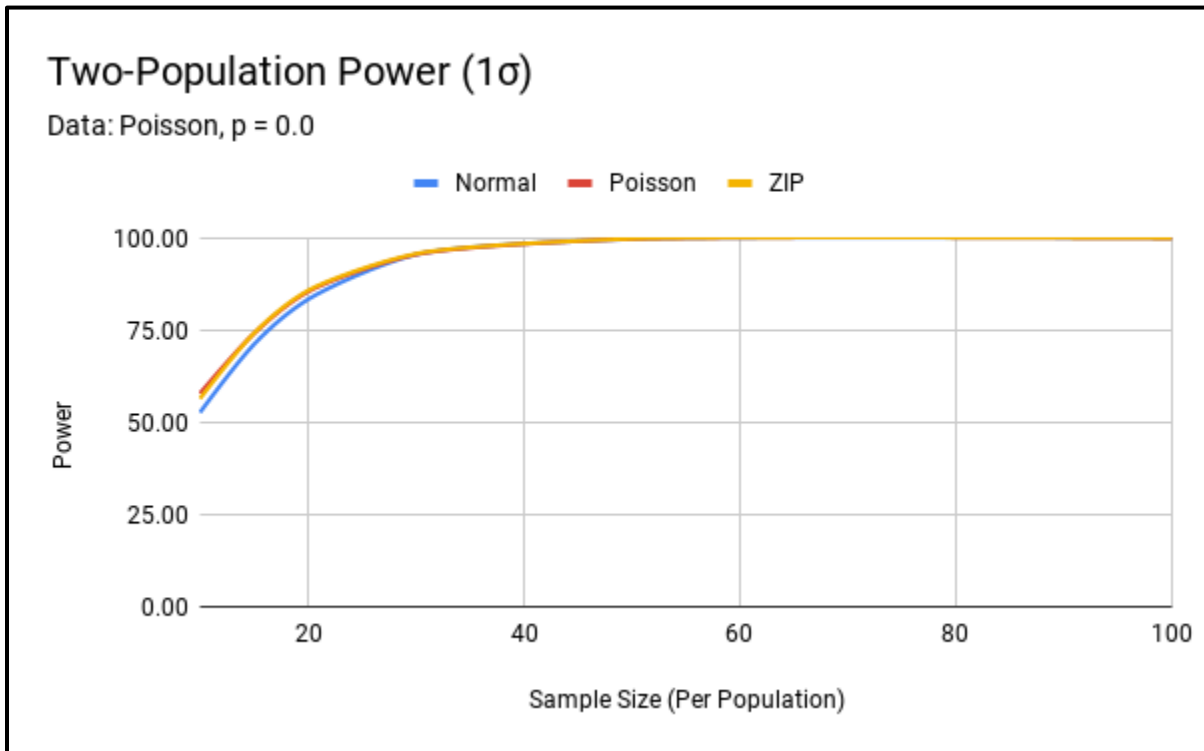


Figure 16. Poisson Sample – $p = 0.0$ Two-Population Power (1σ).

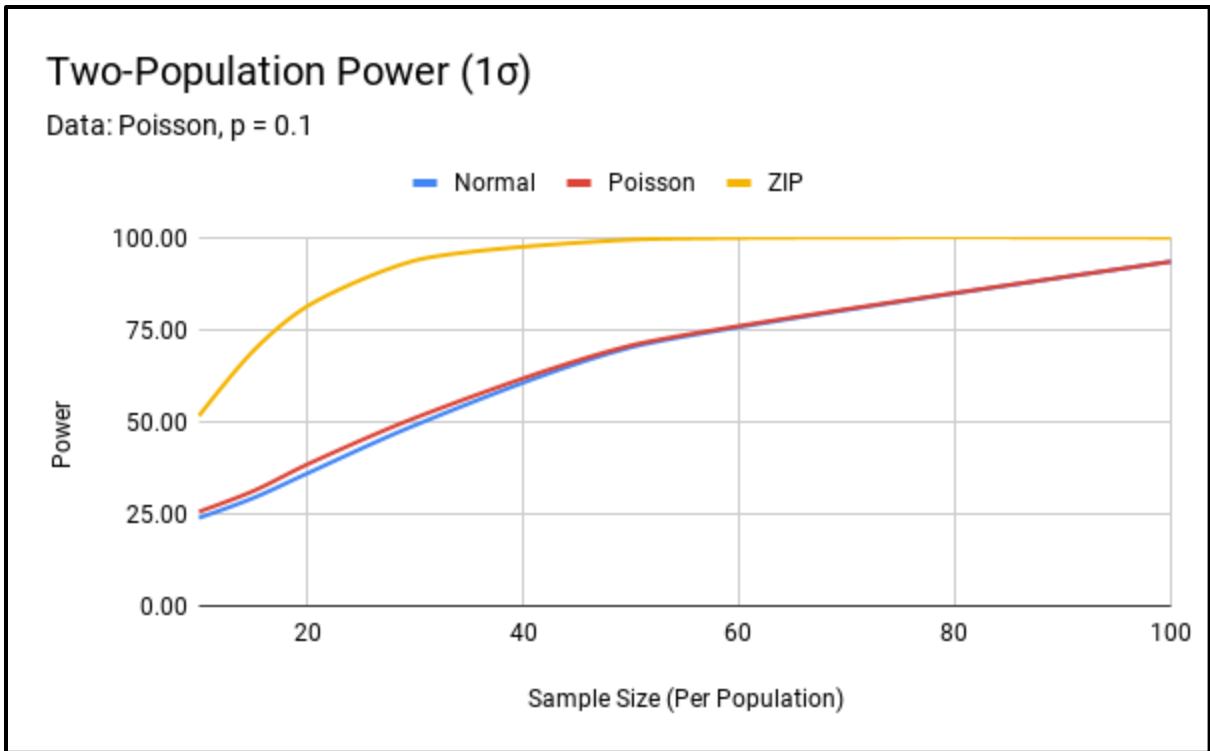


Figure 17. Poisson Sample – $p = 0.1$ Two-Population Power (1σ).

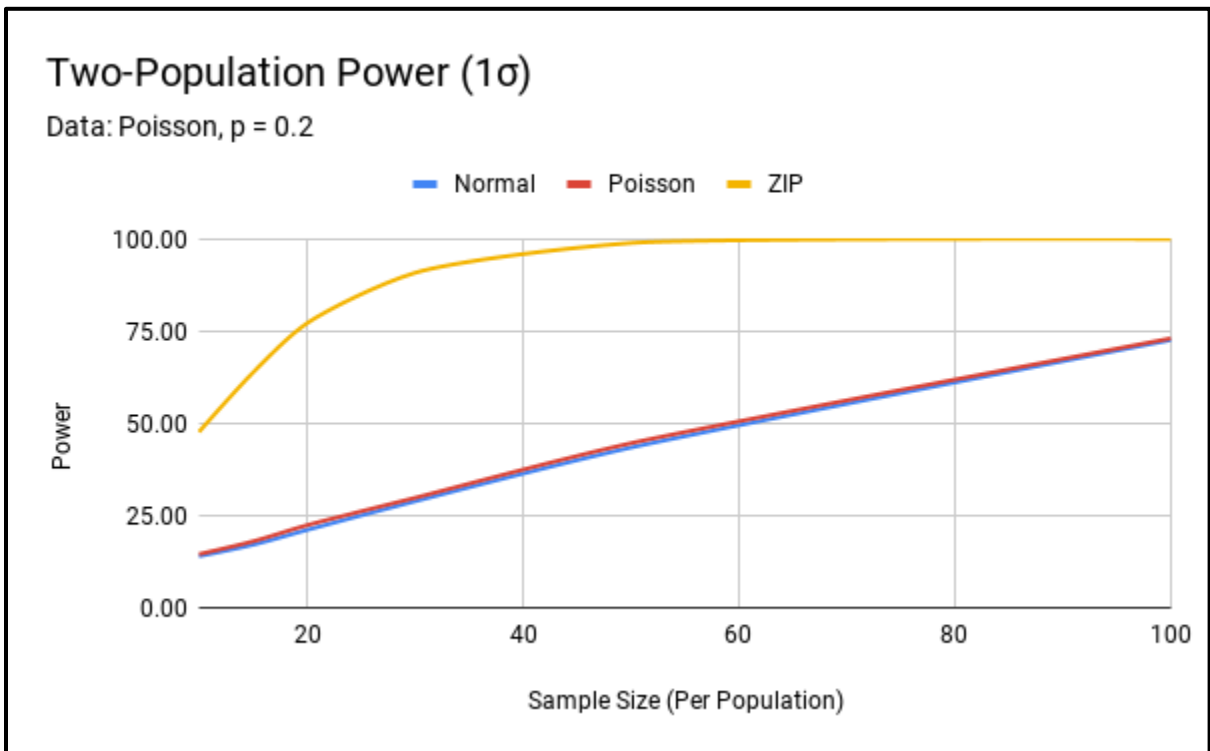


Figure 18. Poisson Sample – $p = 0.2$ Two-Population Power (1σ).

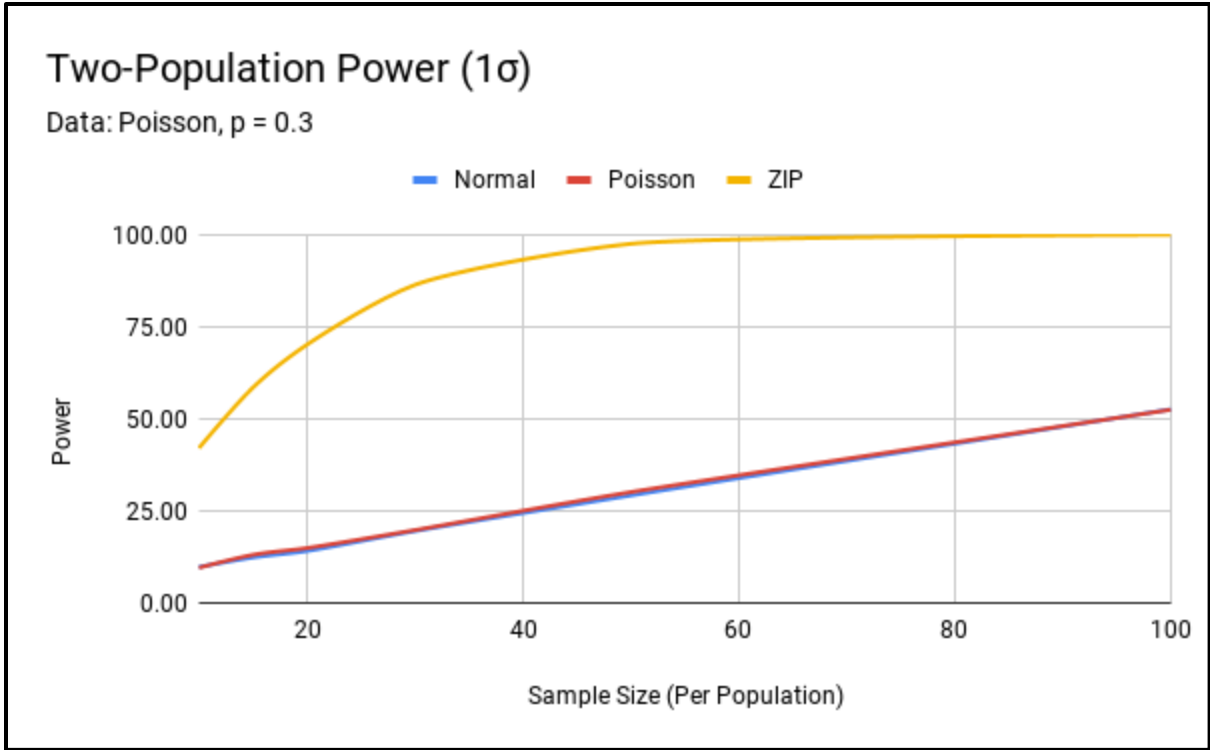


Figure 19. Poisson Sample – $p = 0.3$ Two-Population Power (1σ).

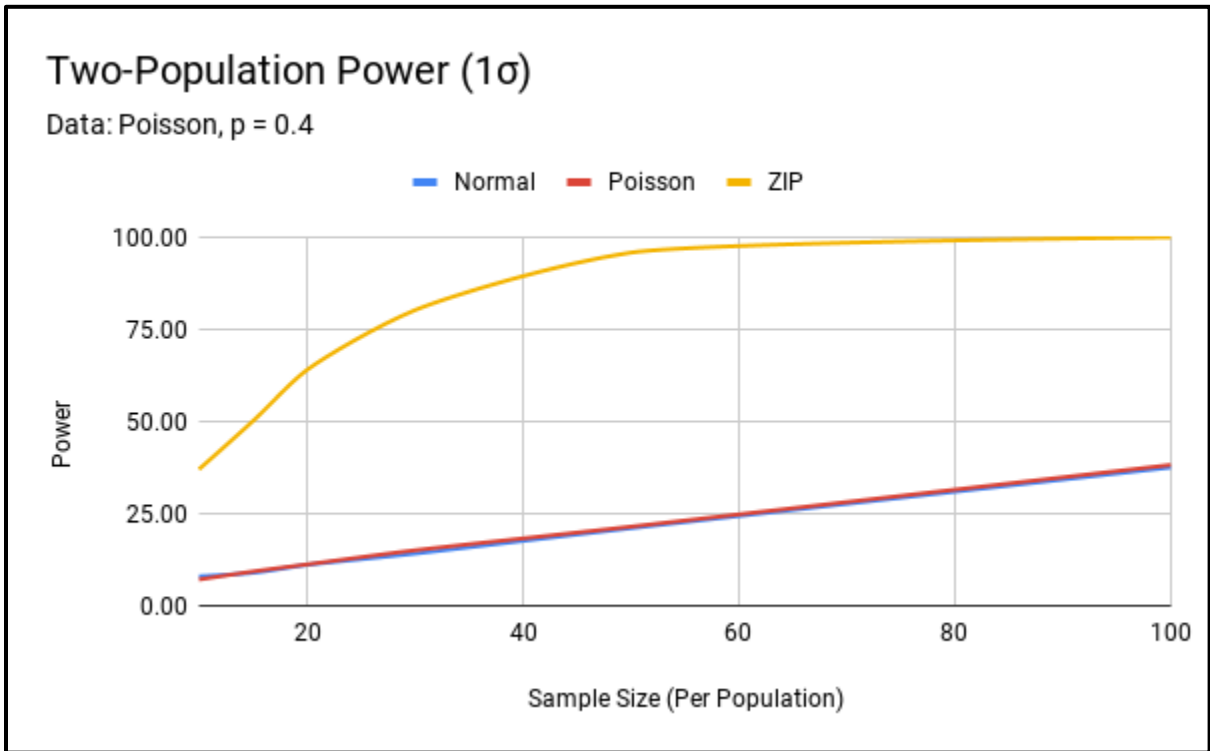


Figure 20. Poisson Sample – $p = 0.4$ Two-Population Power (1σ).

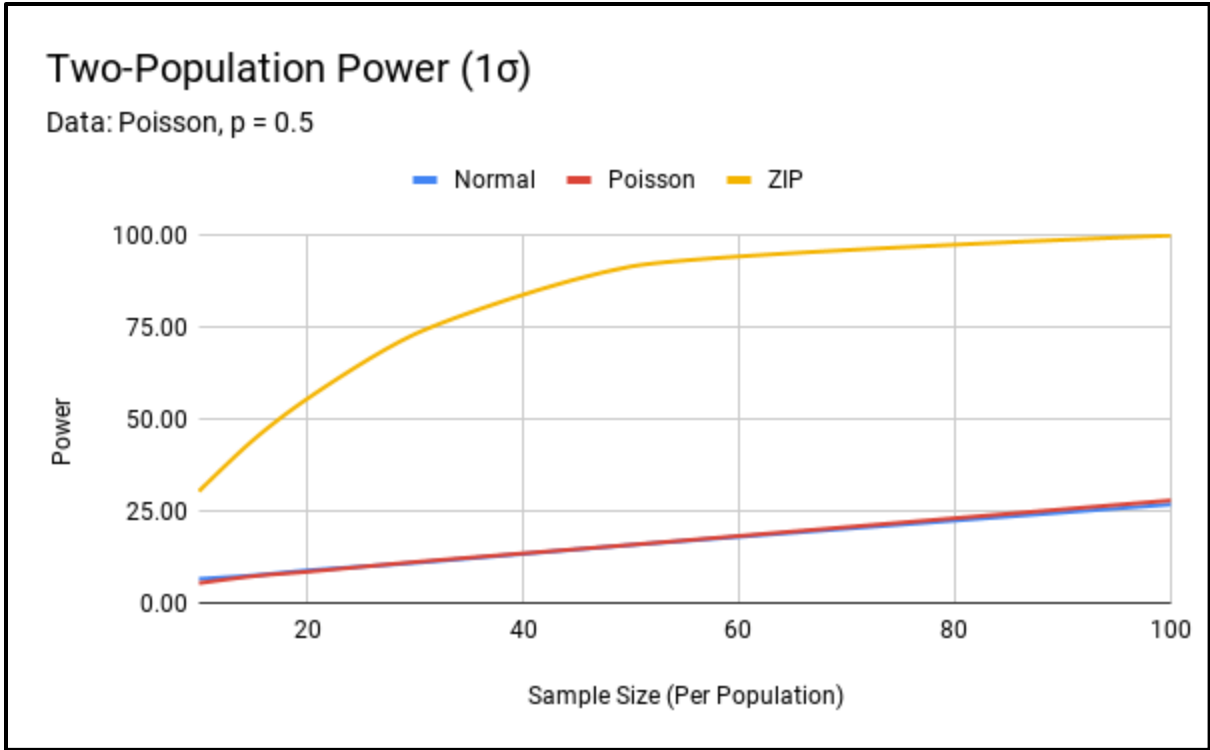


Figure 21. Poisson Sample – $p = 0.5$ Two-Population Power (1σ).

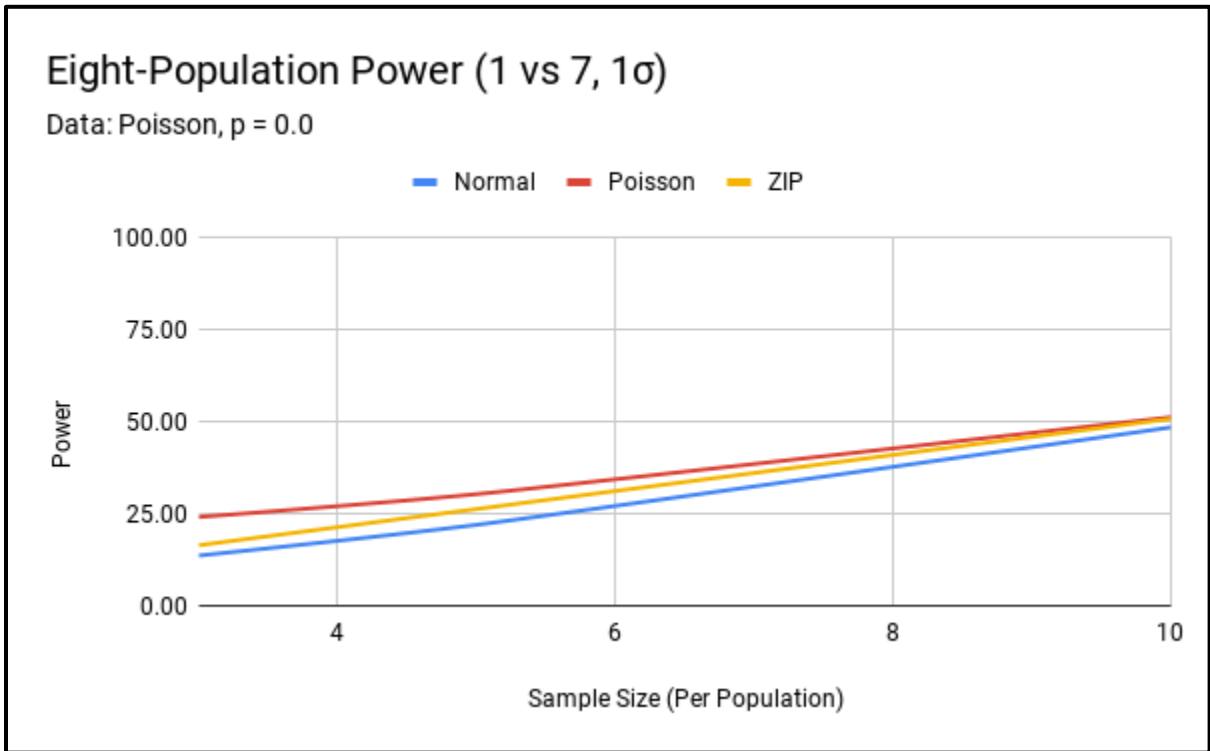


Figure 22. Poisson Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ).

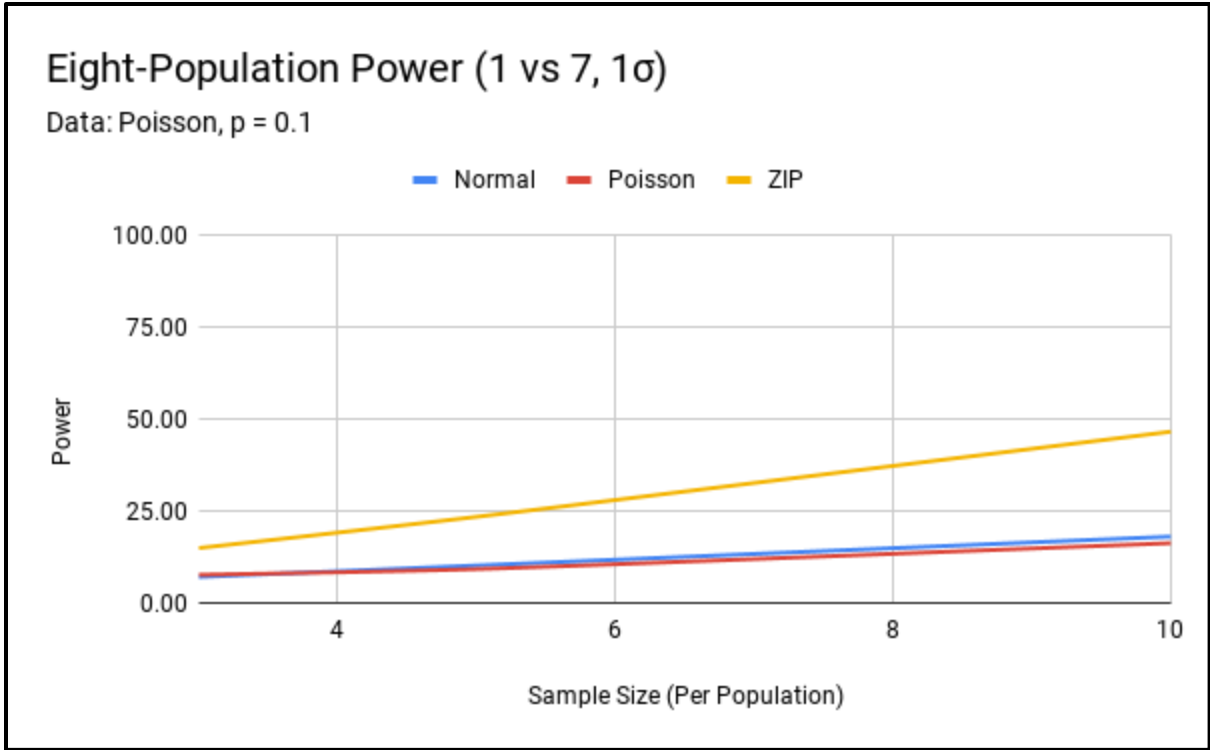


Figure 23. Poisson Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1σ).

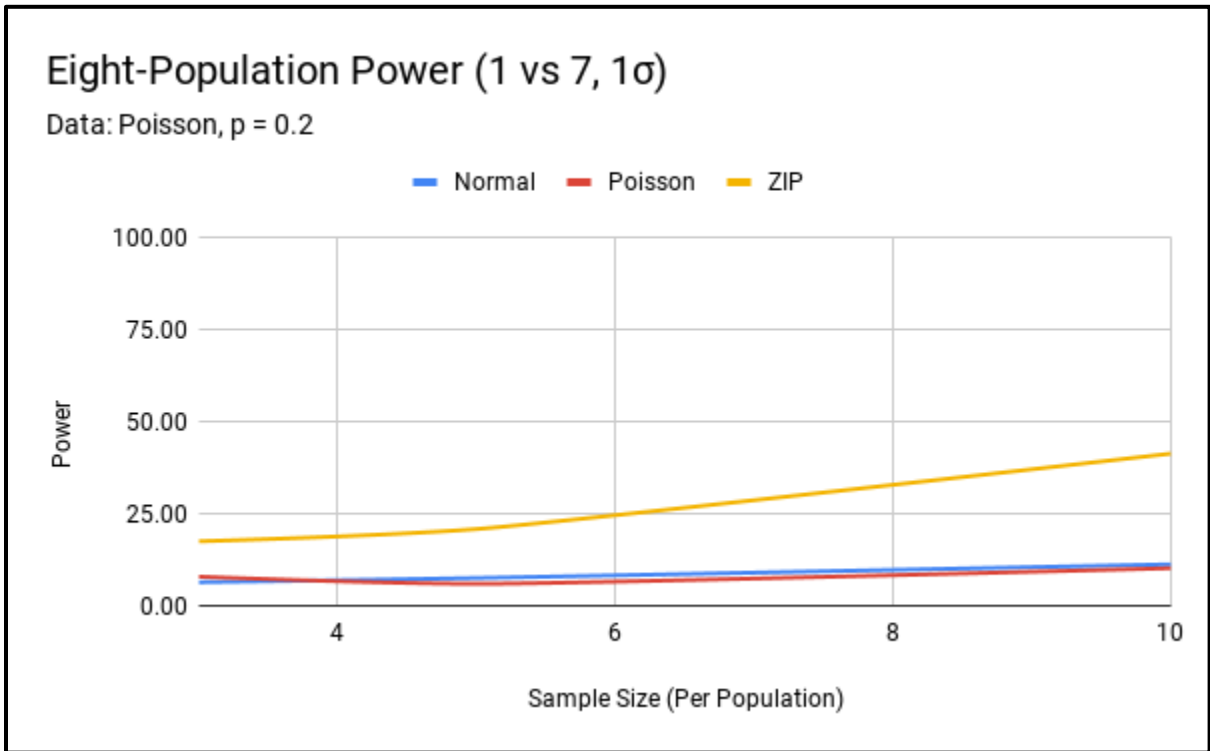


Figure 24. Poisson Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1σ).

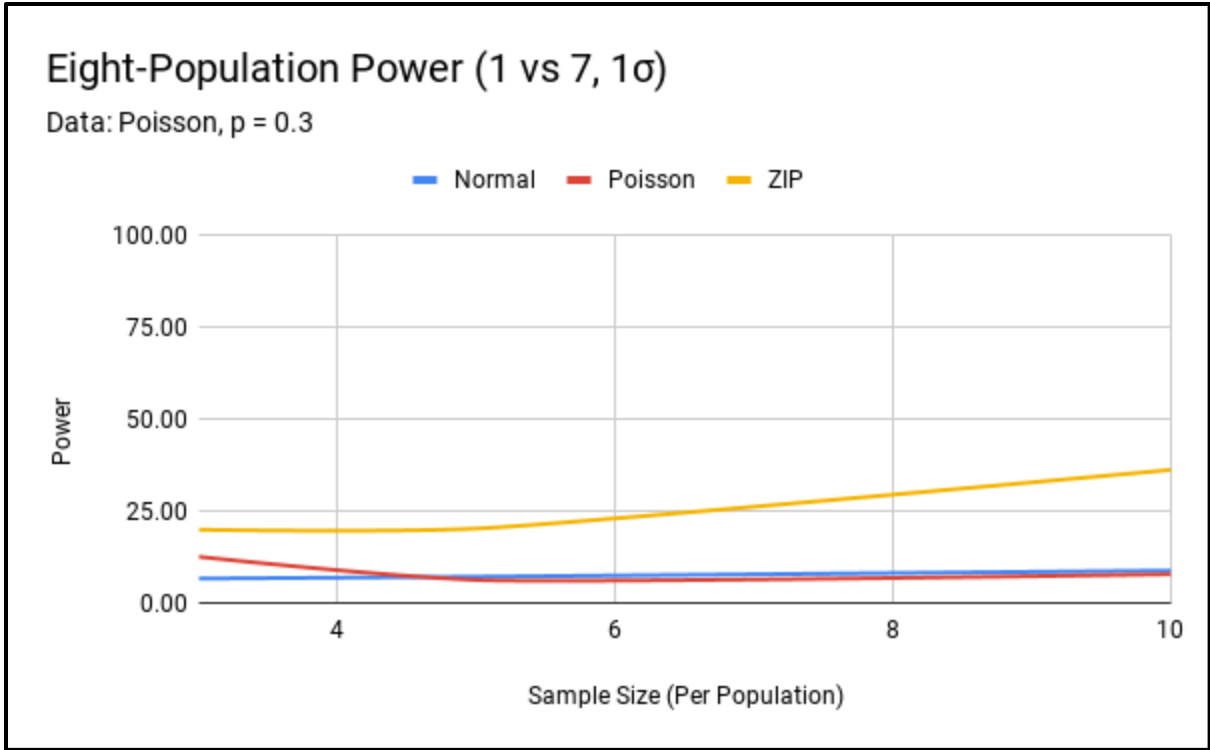


Figure 25. Poisson Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1σ).

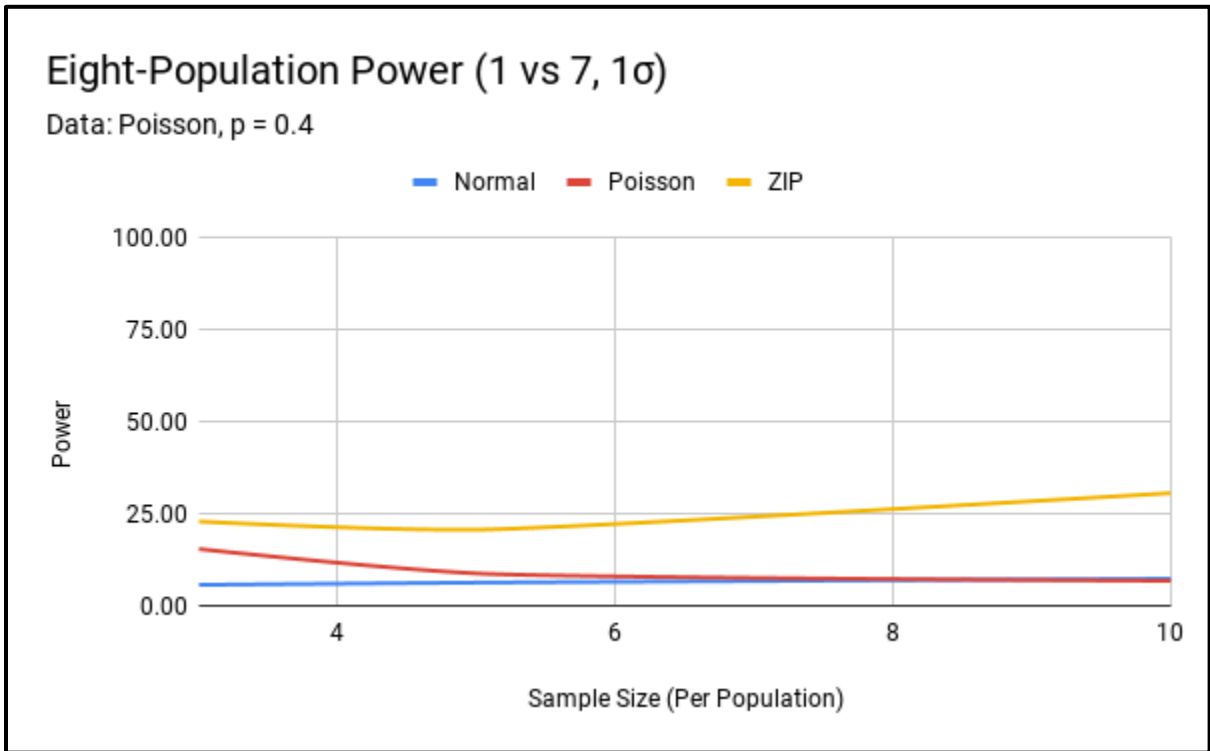


Figure 26. Poisson Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1σ).

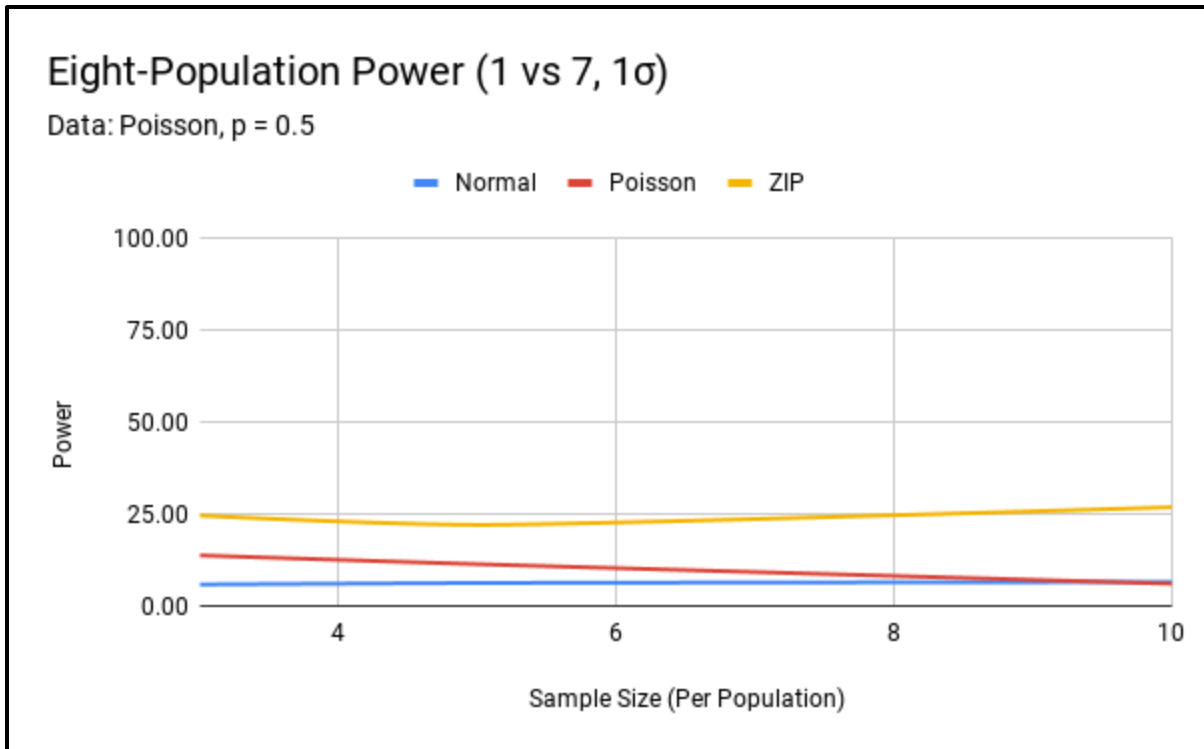


Figure 27. Poisson Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ).

Zero-Inflated Negative Binomial Power Comparison

Due to the inflated Type I error exhibited by the ZIP model applied to zero-inflated negative binomial samples, this model will not be compared to the normal and Poisson models. As seen in the zero-inflated normal and zero-inflated Poisson cases, the normal and Poisson models applied to zero-inflated negative binomial samples provide similar power. The two-population 1σ effect size power comparison plots for the normal and Poisson models can be found in Figures 28, 29, 30, 31, 32, and 33 below. At the 2σ effect size, power between the normal and Poisson models was also similar. To review power plots for the 2σ effect size, see Appendix B.

For eight-population zero-inflated Poisson samples, we only saw acceptable Type I error for samples of $n = 10$ on the normal and Poisson models. Subsequently, we will only compare

those models at the at the $n = 10$ sample size. We saw consistent results between the normal and Poisson models. This result was consistent between the 1 population (1 vs 7) and 4 population (4 vs 4) treatment scenarios. Power comparison plots for the 1 population treatment samples can be found in Figures 34, 35, 36, 37, 38, and 39 below. To review power plots for all combinations of effect size and treatment group scenario, see Appendix B.

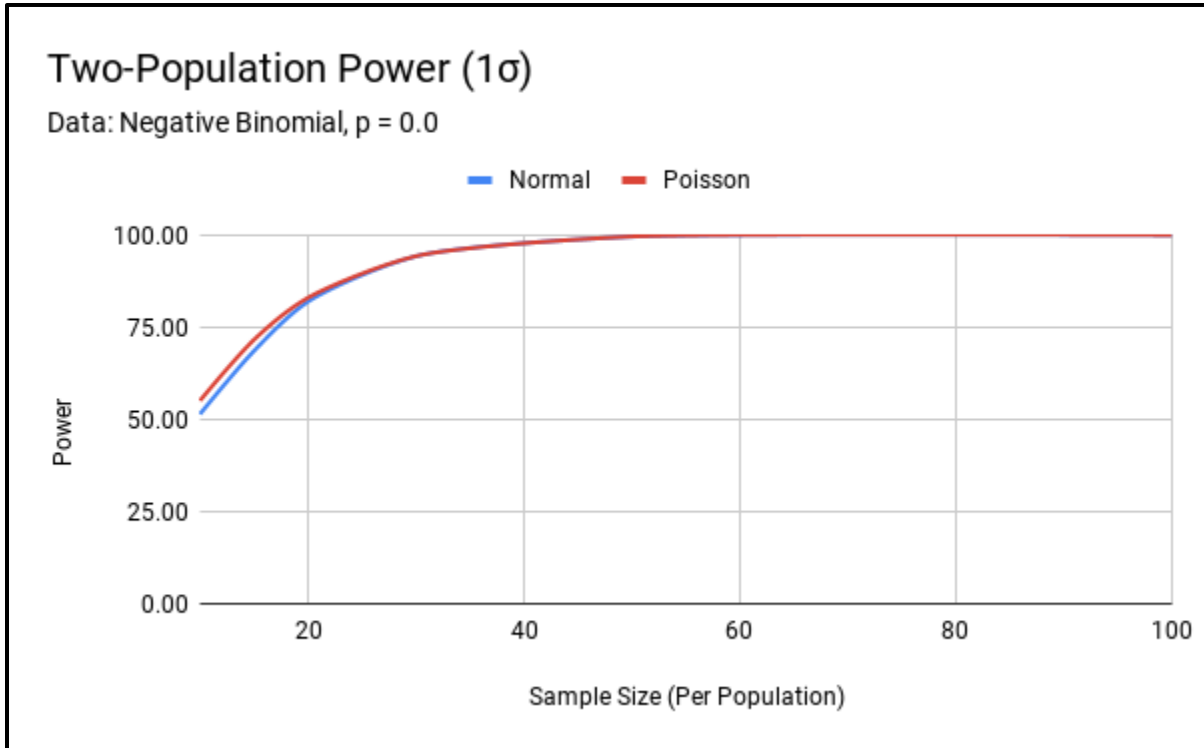


Figure 28. Negative Binomial Sample – $p = 0.0$ Two-Population Power (1σ).

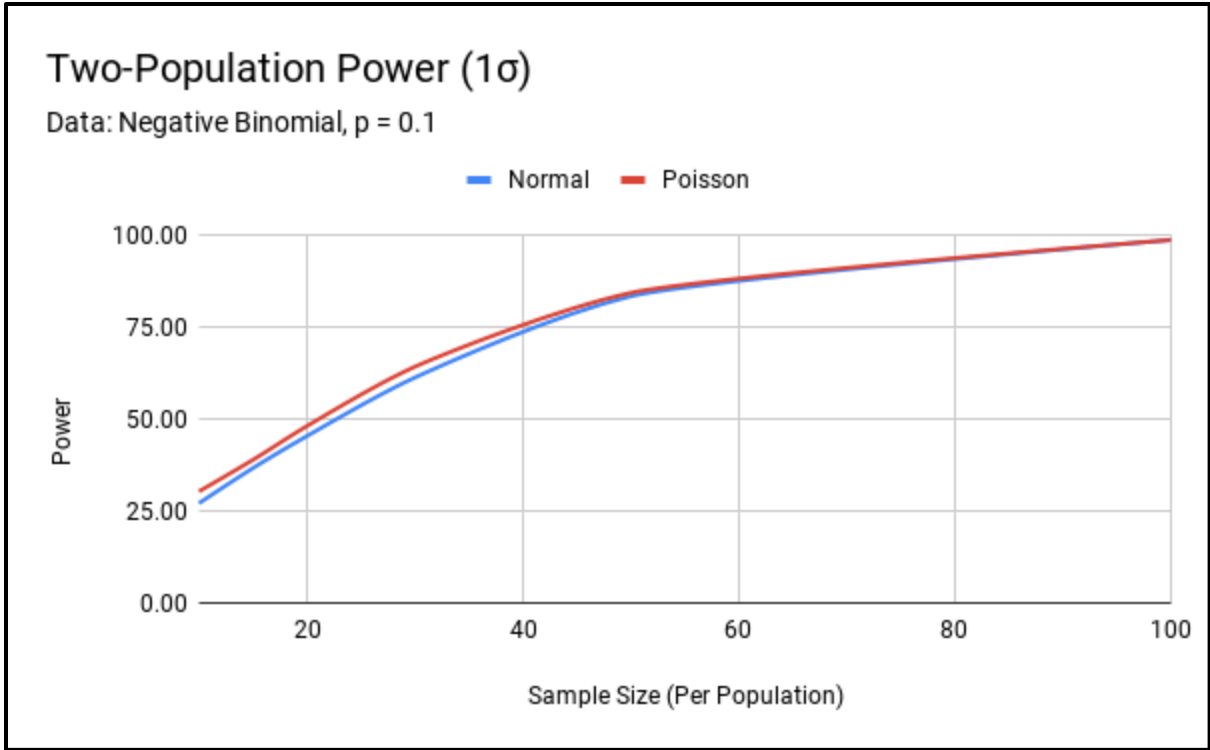


Figure 29. Negative Binomial Sample – $p = 0.1$ Two-Population Power (1σ).

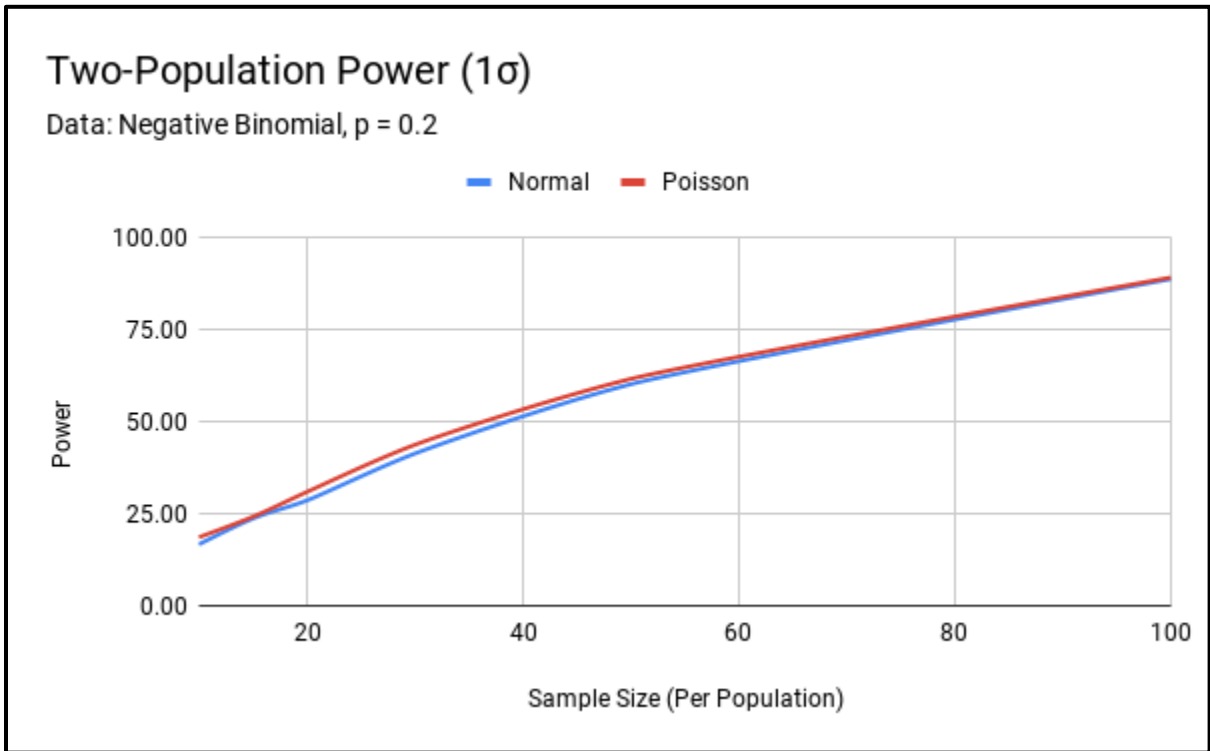


Figure 30. Negative Binomial Sample – $p = 0.2$ Two-Population Power (1σ).

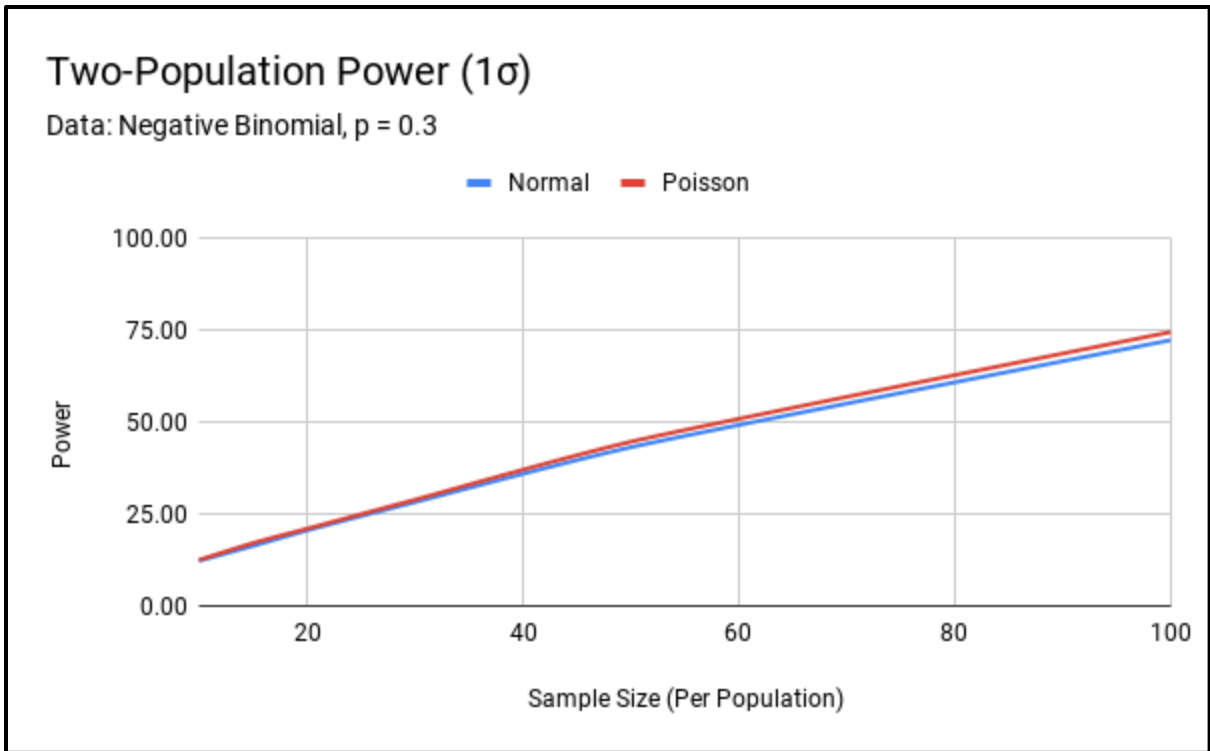


Figure 31. Negative Binomial Sample – $p = 0.3$ Two-Population Power (1σ).

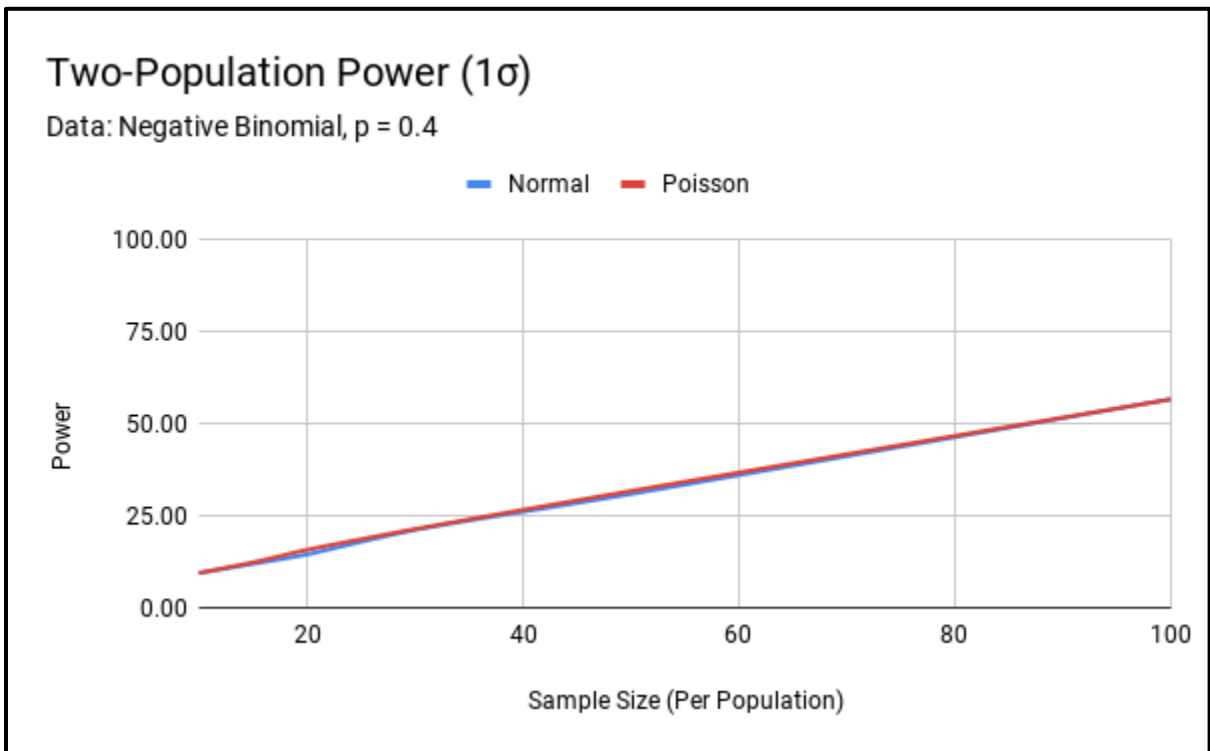


Figure 32. Negative Binomial Sample – $p = 0.4$ Two-Population Power (1σ).

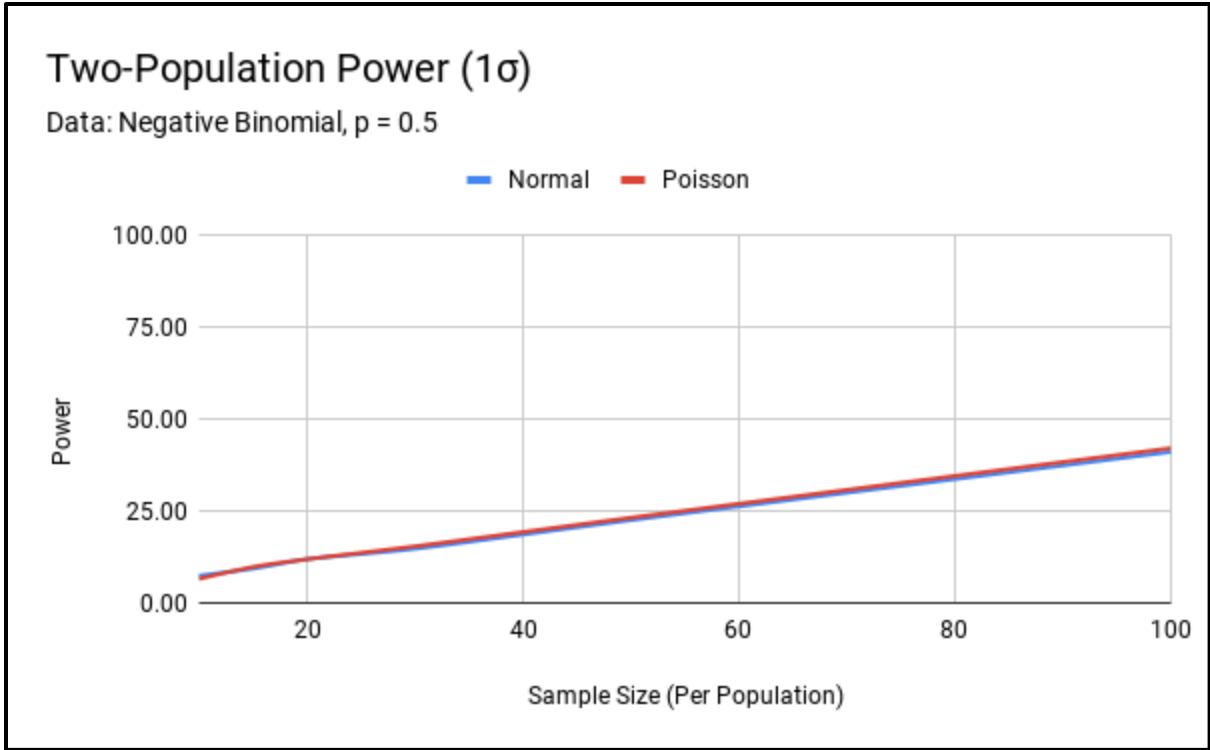


Figure 33. Negative Binomial Sample – $p = 0.5$ Two-Population Power (1σ).

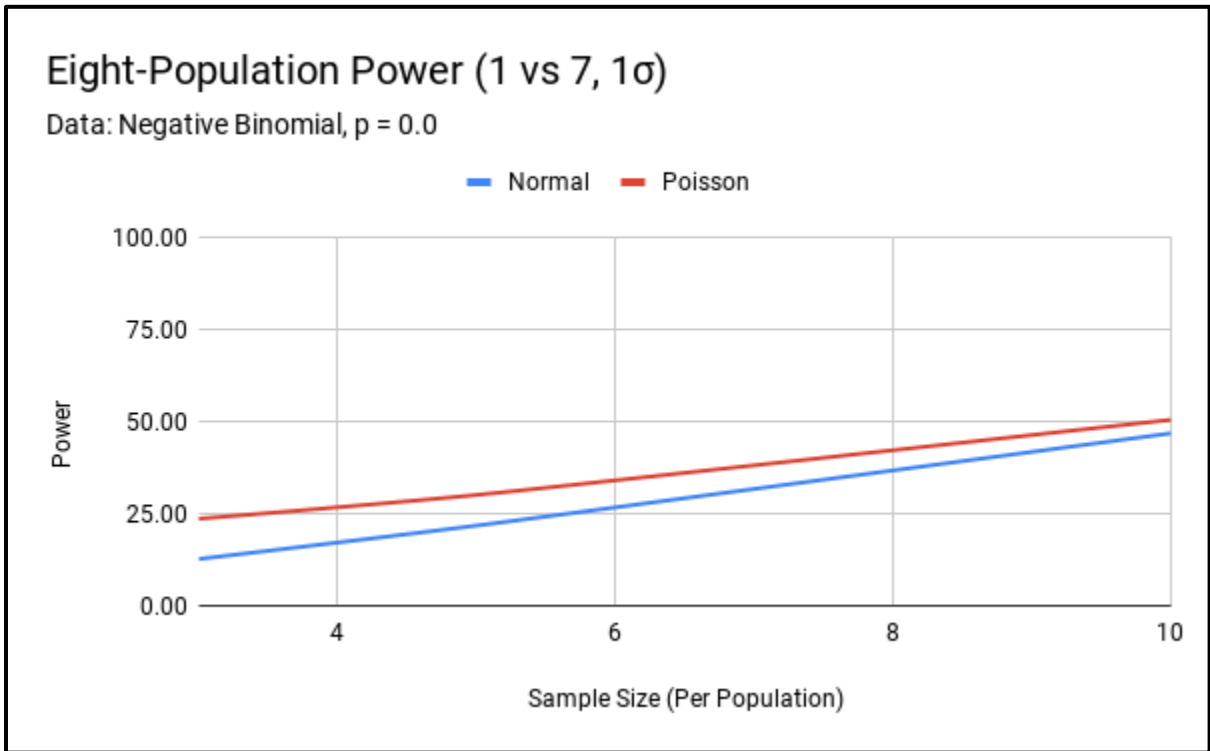


Figure 34. Negative Binomial Sample – $p = 0.0$ Eight-Population Power (1 vs 7, 1σ).

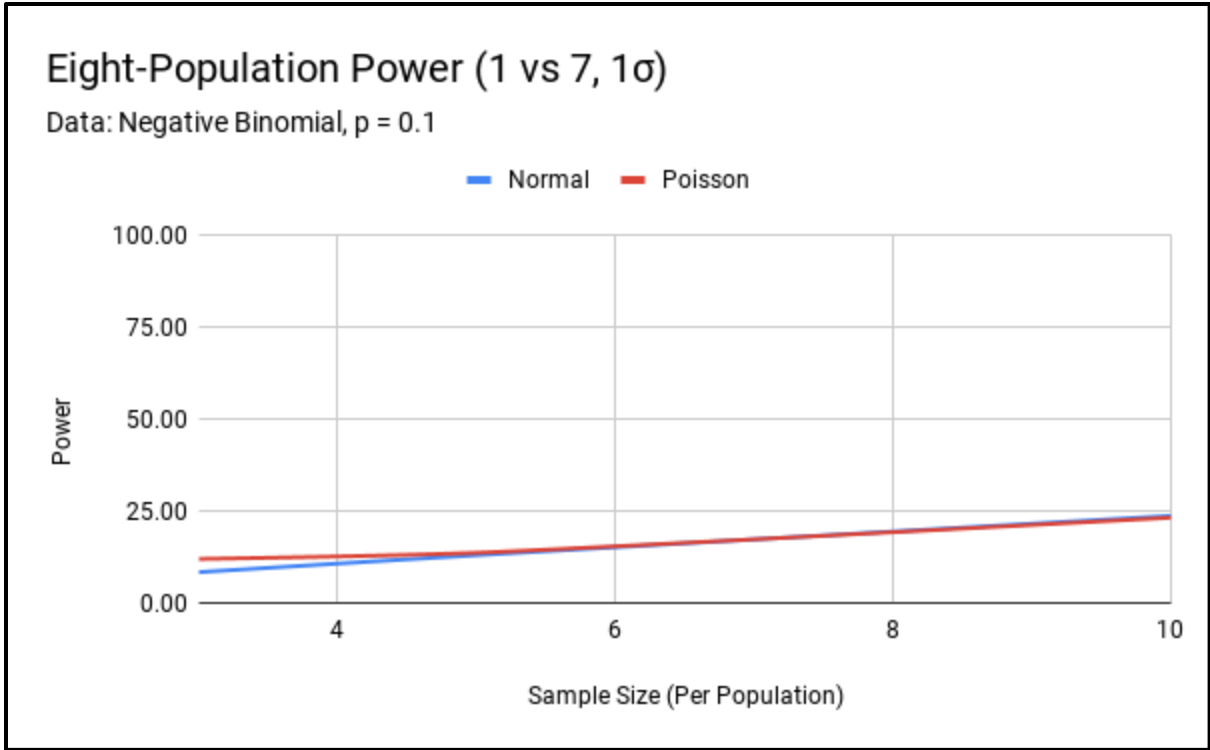


Figure 35. Negative Binomial Sample – $p = 0.1$ Eight-Population Power (1 vs 7, 1 σ).

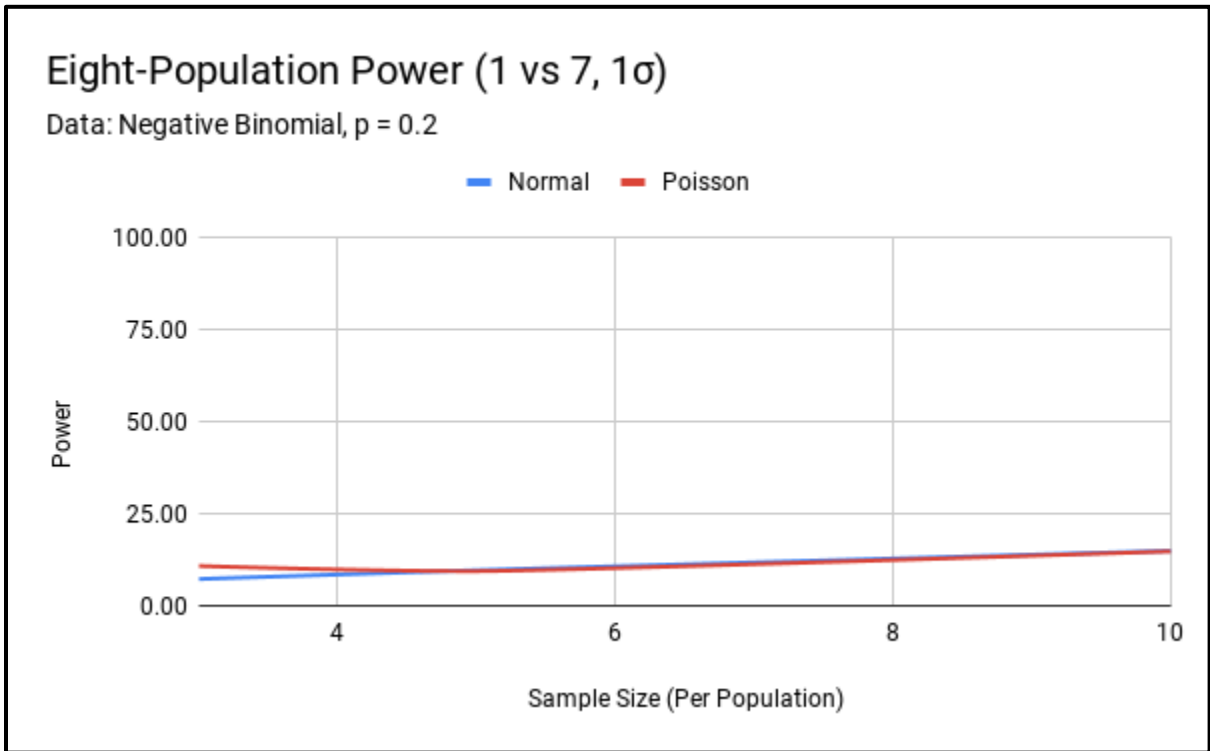


Figure 36. Negative Binomial Sample – $p = 0.2$ Eight-Population Power (1 vs 7, 1 σ).

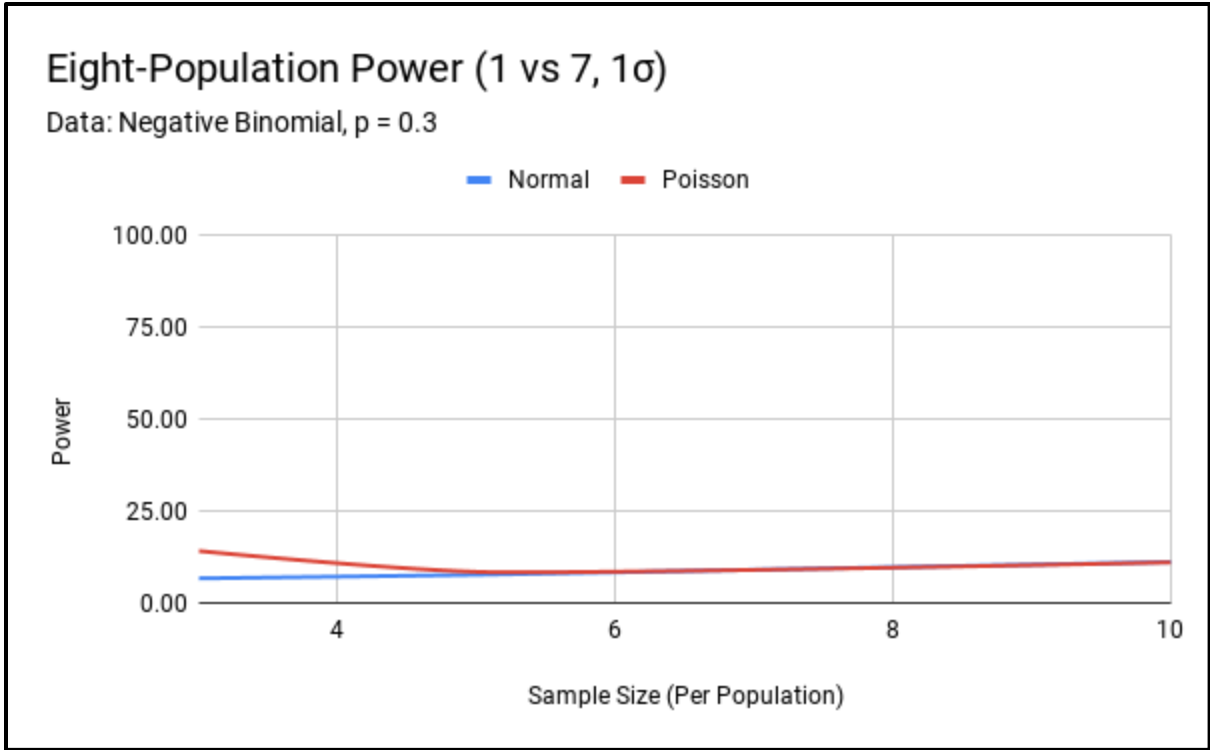


Figure 37. Negative Binomial Sample – $p = 0.3$ Eight-Population Power (1 vs 7, 1 σ).

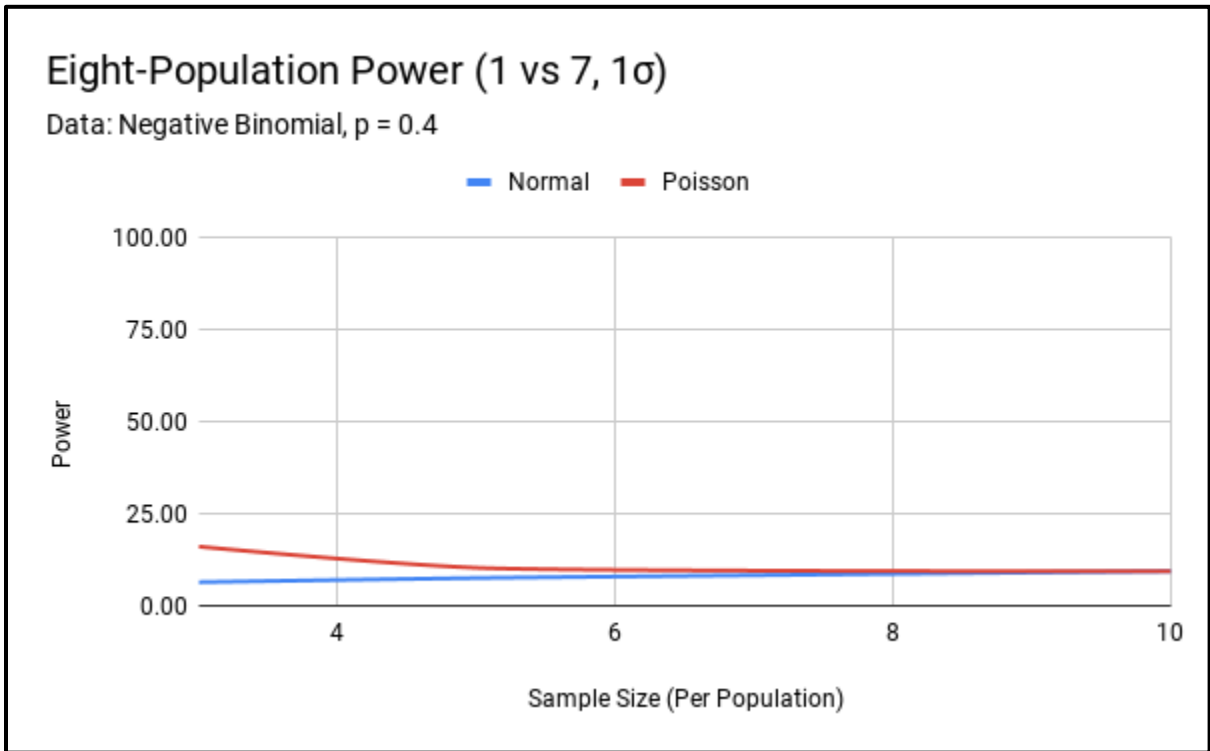


Figure 38. Negative Binomial Sample – $p = 0.4$ Eight-Population Power (1 vs 7, 1 σ).

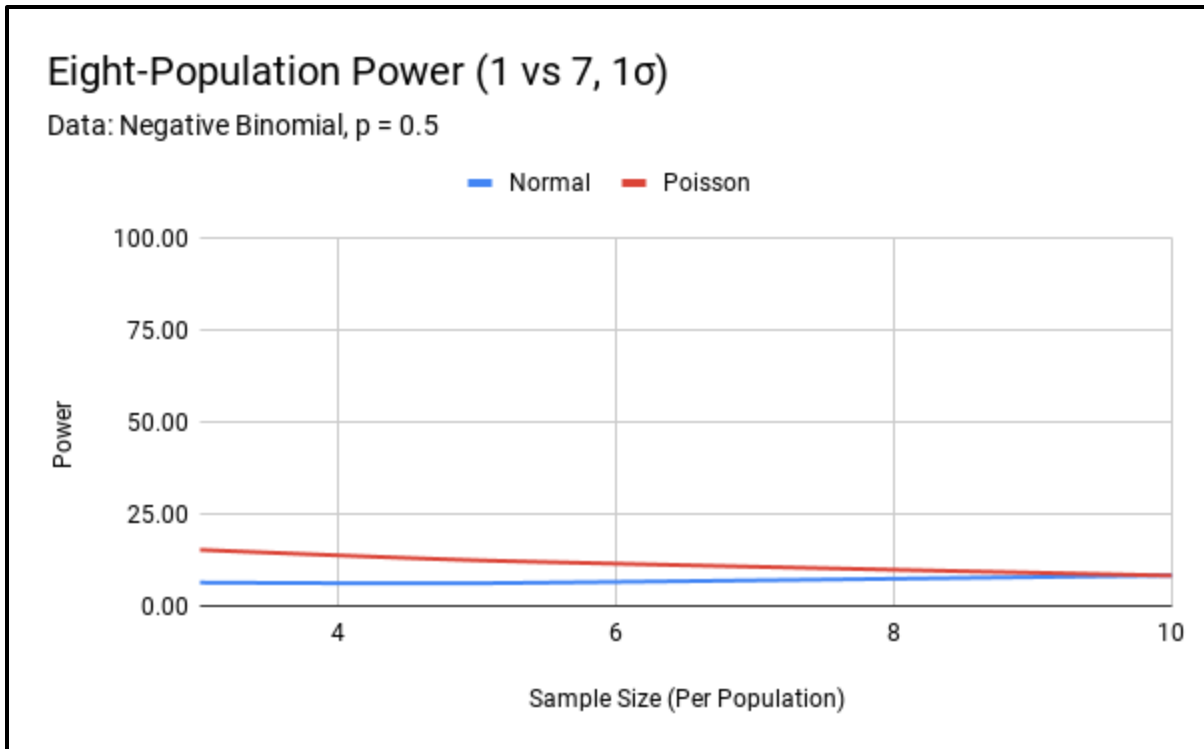


Figure 39. Negative Binomial Sample – $p = 0.5$ Eight-Population Power (1 vs 7, 1σ).

Research Questions

Let's review the research questions proposed in the introduction to this study:

- 1) How does Type I error compare between models specifically designed for zero-inflated data and standard ANOVA?
- 2) Do models specifically designed for zero-inflated data exhibit improved power over ANOVA across different sample sizes?
- 3) What level of zero-inflation significantly affects the Type I error and power of ANOVA?

For the normal model applied to normal samples, Type I error is well controlled. For the Poisson model applied to normal samples with sampling effort of $n = 15$ or greater, Type I error is also well controlled. The ZIP model had low Type I error that was consistently near zero for normal

samples. For Poisson samples, we saw controlled Type I error for all models. For negative binomial samples, we had strong Type I error control for the normal and Poisson models with sampling effort above $n = 10$. We saw that the ZIP model applied to negative binomial samples exhibited high Type I error. To more specifically answer the Type I error research question, for the models and data distributions considered in this study, the zero-inflated method only seems to have better Type I error control when applied to Poisson samples.

Regarding the power of zero-inflated methods when compared to standard ANOVA, in the two-population normal sample treatment scenarios, the ZIP model with samples of $n = 30$ or greater had improved power over the normal or Poisson models. For two-population Poisson samples, the ZIP model showed improved power over the normal or Poisson models for all combinations of sample size and level of zero inflation. In general, these results held true for the eight-population test cases. The notable exception being when we applied the ZIP model to eight-population normal samples; this resulted in extremely low power for all test cases. To review all the power simulations run for this study, see Appendix B.

To address the question on what level of zero-inflation compromises the accuracy of ANOVA, we will review Type I error and power for the normal model. We did not see significant changes in Type I error control for the normal model across levels of zero-inflation for any of our sample distributions (normal, Poisson, and negative binomial). Regarding power, the normal model becomes significantly compromised with nearly any level of zero inflation for all considered sample distributions. We found that the decrease in power was larger for the two-population scenarios than the eight-population scenarios, but an immediate and considerable decrease in power was discovered with all levels of zero-inflation. To review all power findings, see Appendix B.

CHAPTER 5. CONCLUSION

Many tools for the analysis of continuous data have the underlying assumption that the data in question is normal (ex. ANOVA, regression, etc.). Within some disciplines, it is common to end up with a dataset that has a disproportionately high number of zero-values but is otherwise relatively normal. We designed a simulation study to compare the Type I error and power of ANOVA to methods specifically developed for dealing with zero-inflation. In addition to comparing Type I error and power between the models, we considered the following research questions:

- 1) How does Type I error compare between models specifically designed for zero-inflated data and standard ANOVA?
- 2) Do models specifically designed for zero-inflated data exhibit improved power over ANOVA across different sample sizes?
- 3) What level of zero-inflation significantly affects the Type I error and power of ANOVA?

The variables we compared model performance across were sample distribution, level of zero-inflation, sampling effort, and treatment group scenario.

For two-population normal samples, we discovered that the normal and Poisson models could adequately control Type I error. The ZIP model on the other hand exhibited low Type I error, which likely impacted the potential power of that method. The normal and Poisson models exhibited similar power across all levels of zero-inflation and sampling effort. Until reaching sample sizes of $n = 30$, the normal and Poisson models were more powerful for this type of data. For sampling efforts greater than $n = 30$, the ZIP model was more powerful when zero-inflation was present ($p = 0.1$ or greater) despite its low Type I error. For eight-population normal

samples, the ZIP model underperformed the normal and Poisson models across levels of zero-inflation and sampling effort. This was true independent of the treatment group organization (1 population exhibiting treatment effect vs 4 populations exhibiting treatment effect).

For two-population Poisson samples, all three models exhibited relatively strong Type I error control. The normal and Poisson models again had comparable power across levels of zero-inflation and sampling effort. Both the normal and Poisson models were however outperformed by the ZIP model across levels of zero-inflation and sampling effort.

For two-population negative binomial samples, the ZIP model had considerably inflated Type I error and became unsuitable for comparison at any level of sampling effort. The Poisson model only had acceptable Type I error rates with samples of $n = 10$. At that sampling level, the normal and Poisson models have similar performance for the zero-inflated negative binomial samples across levels of zero-inflation. This power consistency between the normal and Poisson models did not diverge for the eight-population samples.

Revisiting our research questions, for the underlying distributions, levels of zero-inflation, sampling efforts, and experimental design scenarios we considered, our models had reasonably strong Type I error control. The most notable exception to that was with the ZIP model applied to negative binomial samples which exhibited high Type I error. For many of the simulation scenarios we considered, the zero-inflated model proved to be more powerful than standard ANOVA when zero-inflation was present in the data. It is worth noting that the zero-inflated methods were not more powerful when there wasn't any zero-inflation. This study also revealed that even low levels of zero-inflation can compromise the power of standard ANOVA for the scenarios we considered.

Overall, ANOVA maintained Type I error for the normal, Poisson, and negative binomial samples with various levels of sample size and zero-inflation. The ZIP model displayed improved power in many situations where the Type I error rate was maintained near the selected alpha level of 0.05. These situations included two-population normal samples with $n = 30$ or greater, and two or eight population Poisson samples with $n = 10$ (specifically with large levels of zero-inflation).

We have proven that using zero-inflated methods on zero-inflated data can be advantageous for increasing statistical power in some situations. In this study, we saw more improvement from the zero-inflated methods for two-population experimental design scenarios than eight-population experimental design scenarios. We have demonstrated and quantified that using ANOVA for zero-inflated data (as previously may have been done) over a model that accounts for zero-inflation can be hindering within certain contexts.

Future Considerations

This study could be expanded by exploring other zero-inflated distributions, the impact of non-constant variance between samples, and other treatment group scenarios. We strictly looked at samples that had equal levels of zero-inflation, so the results from different levels of zero-inflation between samples ($p_1 = 0.1$, $p_2 = 0.3$) could also be explored. We also only considered uniform sampling effort, leaving opportunity to explore the effect of inconsistent sampling effort ($n_1 = 10$, $n_2 = 20$). Additional effort on how to handle the reduction in normalized effect size as zero-inflation increases could provide valuable insight. With additional modeling effort, it is possible that hurdle or Tobit models could also prove to be of value in the realm of zero-inflated data analysis. Further effort to understand why the two-population experimental design scenarios

performed so much better than the eight-population experimental design scenarios could also be beneficial.

REFERENCES

- Amemiya, T. (1973) Regression Analysis when the Dependent Variable Is Truncated Normal. *Econometrica* Vol. 41, No. 6, pp. 997-1016.
- Eggers, J. (2015). On Statistical Methods for Zero-Inflated Models. Technical Report U.U.D.M. Project Report 2015:9, Uppsala University.
- Fisher, R. (1963) Statistical Methods for Research Workers, 13th ed. Edinburgh: Oliver and Boyd.
- Greene, W. (1994) Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. Stern School of Business working paper no. 94-10.
- Hall, D. & Zhang, Z (2004) Marginal models for zero inflated clustered data. *Statistical Modeling* Vol. 4, pp. 161-180.
- Hu, M., Pavlicovia, M. & Nunes, E. (2011) Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse* Vol. 37, No. 5, pp. 367-375.
- Lambert, D. (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* Vol. 34, No. 1, pp. 1-14.
- Lord, D., Washington, S., & Ivan, J. (2005) Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention* Vol. 37, No. 1, pp. 35-46.
- McDonald, F. & Moffitt, R. (1980) The Uses of Tobit Analysis. *The Review of Economics and Statistics* Vol. 62, No. 2, pp. 318-321.
- Min, Y. & Agresti, A. (2002) Modeling Nonnegative Data with Clumping at Zero: A Survey. *Journal of Iranian Statistical Society* Vol. 1, pp. 7-33.

- Mullahy, J. (1986) Specification of Some Modified Count Data Models. *Journal of Econometrics* Vol. 33, pp. 341-365.
- Sileshi, G. (2008) The Excess-Zero Problem in Soil Animal Count Data and Choice of Appropriate Models for Statistical Inference. *Pedobiologia* Vol. 52, pp. 1-17.
- Stroup, W. (2013) Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. CRC Press.
- Tobin, J. (1958) Estimation of Relationships for Limited Dependent Variables. *Econometrica* Vol. 26, No. 1, pp. 24-36.
- Yang, S., Harlow, L., Puggioni, G., & Redding, C. (2017) A Comparison of Different Methods of Zero-Inflated Data Analysis and an Application in Health. *Journal of Modern Applied Statistical Methods* Vol 16, No. 1, pp. 518-543.

APPENDIX A. TYPE I ERROR ASSESSMENT RESULTS

Two-Population Normal Samples

Sample distribution: Normal

Simulations: 10,000

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = \mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

Table A1. Normal Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.95 | 2.82 | 4.78 | 4.92 | 4.98 | 5.06 |
| n = 15 | 4.94 | 4.09 | 4.63 | 5.55 | 4.99 | 5.29 |
| n = 20 | 4.95 | 4.52 | 4.94 | 4.81 | 5.10 | 4.87 |
| n = 30 | 5.03 | 4.36 | 4.70 | 5.30 | 4.71 | 5.10 |
| n = 50 | 4.94 | 4.65 | 5.11 | 4.93 | 4.88 | 5.23 |
| n = 100 | 5.05 | 5.02 | 4.95 | 4.74 | 5.14 | 4.90 |

Table A2. Normal Sample – Poisson Model Two-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.54 | 2.43 | 3.51 | 3.84 | 3.72 | 3.29 |
| n = 15 | 5.64 | 3.64 | 4.53 | 4.45 | 3.80 | 3.76 |
| n = 20 | 6.03 | 4.10 | 4.58 | 4.70 | 4.70 | 4.20 |
| n = 30 | 5.50 | 4.34 | 4.47 | 4.77 | 4.49 | 4.59 |
| n = 50 | 5.47 | 4.23 | 4.77 | 4.43 | 5.09 | 4.87 |
| n = 100 | 5.21 | 5.00 | 4.65 | 4.83 | 5.04 | 5.42 |

Table A3. Normal Sample – ZIP Model Two-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| n = 20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| n = 30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| n = 100 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |

Two-Population Poisson Samples

Sample distribution: Poisson

Simulations: 10,000

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = \lambda_2 = 20$

Table A4. Poisson Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 5.07 | 4.19 | 4.85 | 4.97 | 5.02 | 4.97 |
| n = 15 | 5.02 | 4.86 | 5.42 | 4.81 | 5.14 | 4.88 |
| n = 20 | 4.97 | 4.82 | 4.88 | 5.16 | 4.77 | 5.13 |
| n = 30 | 5.07 | 5.14 | 5.21 | 5.26 | 5.20 | 5.01 |
| n = 50 | 4.57 | 5.22 | 5.12 | 5.19 | 5.06 | 5.02 |
| n = 100 | 5.10 | 5.40 | 4.84 | 5.10 | 4.97 | 4.96 |

Table A5. Poisson Sample – Poisson Model Two-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.42 | 4.80 | 4.45 | 4.15 | 4.10 | 3.53 |
| n = 15 | 5.96 | 5.10 | 4.96 | 4.15 | 4.02 | 4.16 |
| n = 20 | 5.73 | 4.61 | 4.78 | 4.87 | 4.62 | 4.49 |
| n = 30 | 5.54 | 5.08 | 5.25 | 4.75 | 4.84 | 4.22 |
| n = 50 | 5.02 | 4.74 | 4.93 | 4.80 | 4.91 | 5.01 |
| n = 100 | 5.31 | 5.18 | 5.00 | 4.81 | 4.42 | 4.76 |

Table A6. Poisson Sample – ZIP Model Two-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.86 | 4.73 | 4.95 | 4.85 | 4.92 | 4.57 |
| n = 15 | 4.78 | 5.07 | 4.87 | 5.11 | 4.71 | 4.85 |
| n = 20 | 5.00 | 4.67 | 4.89 | 5.11 | 4.78 | 5.26 |
| n = 30 | 5.04 | 4.89 | 4.95 | 4.68 | 4.64 | 5.00 |
| n = 50 | 4.77 | 4.82 | 4.95 | 4.78 | 5.04 | 4.92 |
| n = 100 | 5.15 | 5.35 | 4.89 | 5.08 | 4.87 | 4.71 |

Two-Population Negative Binomial Samples

Sample distribution: Negative Binomial

Simulations: 10,000

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

Table A7. Negative Binomial Sample – Normal Model Two-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 4.80 | 5.48 | 4.80 | 5.58 | 4.82 | 4.88 |
| n = 15 | 5.15 | 4.66 | 4.85 | 4.64 | 4.74 | 5.13 |
| n = 20 | 4.74 | 5.13 | 4.82 | 5.26 | 5.17 | 4.99 |
| n = 30 | 4.84 | 5.27 | 4.64 | 5.25 | 4.84 | 5.18 |
| n = 50 | 5.13 | 4.82 | 4.86 | 5.16 | 4.99 | 4.94 |
| n = 100 | 5.06 | 4.98 | 4.95 | 4.97 | 4.70 | 5.09 |

Table A8. Negative Binomial Sample – Poisson Model Two-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 6.65 | 5.76 | 5.06 | 4.09 | 4.51 | 3.74 |
| n = 15 | 5.61 | 5.19 | 4.81 | 4.78 | 4.45 | 4.38 |
| n = 20 | 5.66 | 5.31 | 4.87 | 5.25 | 4.54 | 4.81 |
| n = 30 | 5.03 | 5.20 | 5.02 | 4.82 | 4.39 | 4.99 |
| n = 50 | 5.46 | 4.80 | 4.91 | 4.73 | 4.62 | 4.69 |
| n = 100 | 5.42 | 4.76 | 5.03 | 5.18 | 4.92 | 5.00 |

Table A9. Negative Binomial Sample – ZIP Model Two-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 16.75 | 16.42 | 17.18 | 15.81 | 15.69 | 16.43 |
| n = 15 | 16.62 | 16.11 | 16.38 | 16.12 | 16.01 | 16.01 |
| n = 20 | 17.19 | 16.56 | 16.46 | 16.31 | 17.02 | 16.04 |
| n = 30 | 15.69 | 16.21 | 15.99 | 16.05 | 16.35 | 17.23 |
| n = 50 | 17.01 | 16.82 | 16.45 | 16.03 | 17.17 | 16.36 |
| n = 100 | 16.84 | 16.53 | 16.25 | 16.20 | 16.23 | 16.45 |

Eight-Population Normal Samples

Sample distribution: Normal

Simulations: 10,000

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

Table A10. Normal Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 4.89 | 4.33 | 5.12 | 5.21 | 6.01 | 5.62 |
| n = 5 | 5.17 | 4.20 | 5.34 | 4.97 | 5.26 | 5.34 |
| n = 10 | 4.93 | 4.16 | 5.13 | 4.82 | 5.32 | 4.86 |

Table A11. Normal Sample – Poisson Model Eight-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.14 | 2.51 | 5.93 | 12.18 | 14.68 | 10.79 |
| n = 5 | 8.62 | 2.64 | 2.86 | 3.99 | 7.59 | 11.42 |
| n = 10 | 6.33 | 3.58 | 3.78 | 4.06 | 4.39 | 4.23 |

Table A12. Normal Sample – ZIP Model Eight-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 0.00 | 0.17 | 0.40 | 0.63 | 1.24 | 1.28 |
| n = 5 | 0.00 | 0.00 | 0.13 | 0.41 | 0.33 | 0.78 |
| n = 10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.31 |

Eight-Population Poisson Samples

Sample distribution: Poisson

Simulations: 10,000

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

Table A13. Poisson Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 5.22 | 4.71 | 5.11 | 5.74 | 5.30 | 5.19 |
| n = 5 | 4.83 | 4.51 | 4.75 | 4.32 | 5.43 | 5.15 |
| n = 10 | 4.86 | 4.93 | 5.32 | 5.19 | 4.89 | 5.55 |

Table A14. Poisson Sample – Poisson Model Eight-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.02 | 4.57 | 6.96 | 12.32 | 14.18 | 12.24 |
| n = 5 | 8.38 | 3.98 | 4.33 | 4.54 | 7.39 | 10.70 |
| n = 10 | 6.35 | 4.51 | 4.22 | 4.50 | 4.86 | 5.11 |

Table A15. Poisson Sample – ZIP Model Eight-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 5.42 | 5.78 | 8.08 | 11.21 | 13.92 | 16.99 |
| n = 5 | 5.22 | 4.79 | 5.22 | 6.21 | 8.34 | 11.78 |
| n = 10 | 4.47 | 5.36 | 5.10 | 4.97 | 5.23 | 5.88 |

Eight-Population Negative Binomial Samples

Sample distribution: Negative Binomial

Simulations: 10,000

$$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$$

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$$

$$k(\text{nb})_1 = k(\text{nb})_2 = k(\text{nb})_3 = k(\text{nb})_4 = k(\text{nb})_5 = k(\text{nb})_6 = k(\text{nb})_7 = k(\text{nb})_8 = 20$$

$$p(\text{nb})_1 = p(\text{nb})_2 = p(\text{nb})_3 = p(\text{nb})_4 = p(\text{nb})_5 = p(\text{nb})_6 = p(\text{nb})_7 = p(\text{nb})_8 = 0.5$$

Table A16. Negative Binomial Sample – Normal Model Eight-Population Type I Error.

| Zero-Inflation | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 4.75 | 4.82 | 5.28 | 5.26 | 5.46 | 5.33 |
| n = 5 | 4.88 | 4.99 | 5.16 | 4.81 | 4.99 | 5.32 |
| n = 10 | 4.68 | 4.55 | 5.67 | 5.18 | 5.27 | 5.30 |

Table A17. Negative Binomial Sample – Poisson Model Eight-Population Type I Error.

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.41 | 7.01 | 8.69 | 12.82 | 14.67 | 13.20 |
| n = 5 | 8.31 | 5.76 | 4.87 | 5.81 | 8.61 | 11.04 |
| n = 10 | 6.64 | 4.96 | 5.34 | 5.05 | 4.95 | 5.65 |

Table A18. Negative Binomial Sample – ZIP Model Eight-Population Type I Error.

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 42.61 | 42.88 | 45.24 | 49.41 | 52.41 | 55.01 |
| n = 5 | 42.63 | 42.31 | 41.73 | 43.32 | 45.90 | 48.49 |
| n = 10 | 42.50 | 43.12 | 43.43 | 42.35 | 43.38 | 42.98 |

APPENDIX B. POWER COMPARISON RESULTS

Two-Population Scenario

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 22$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 56.03 | 16.85 | 8.70 | 6.67 | 5.84 | 5.68 |
| n = 15 | 74.88 | 17.26 | 8.43 | 6.68 | 6.18 | 5.51 |
| n = 20 | 86.93 | 16.99 | 9.40 | 8.01 | 6.64 | 6.36 |
| n = 30 | 96.67 | 20.79 | 11.55 | 8.23 | 7.50 | 6.86 |
| n = 50 | 99.87 | 28.99 | 16.68 | 10.84 | 8.70 | 7.82 |
| n = 100 | 100.00 | 49.88 | 26.37 | 16.96 | 13.33 | 10.15 |

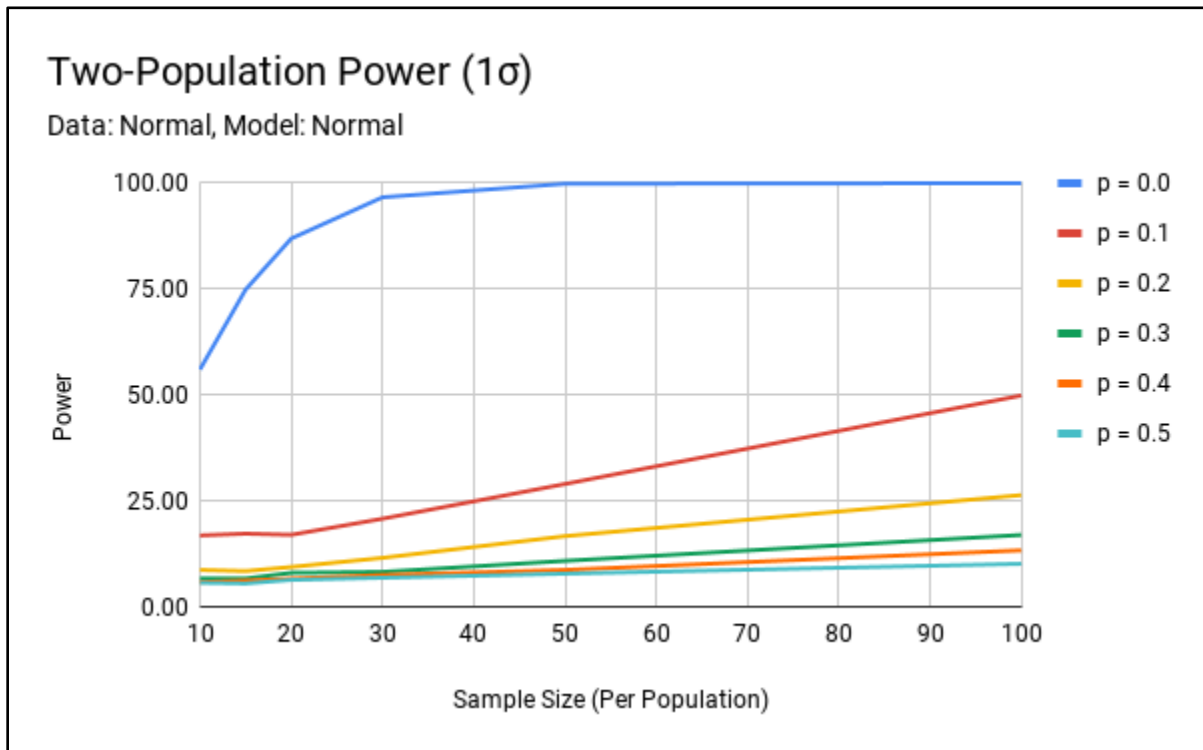


Figure B1. Normal Sample – Normal Model Two-Population Power (1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 22$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 61.50 | 17.28 | 7.83 | 5.64 | 4.25 | 3.68 |
| n = 15 | 78.04 | 17.96 | 8.28 | 6.41 | 5.41 | 4.73 |
| n = 20 | 88.19 | 17.95 | 9.33 | 7.63 | 5.80 | 5.32 |
| n = 30 | 97.02 | 20.53 | 11.69 | 8.59 | 7.31 | 6.22 |
| n = 50 | 99.85 | 29.63 | 15.19 | 11.21 | 9.01 | 7.41 |
| n = 100 | 100.00 | 49.20 | 26.54 | 17.49 | 12.93 | 10.27 |

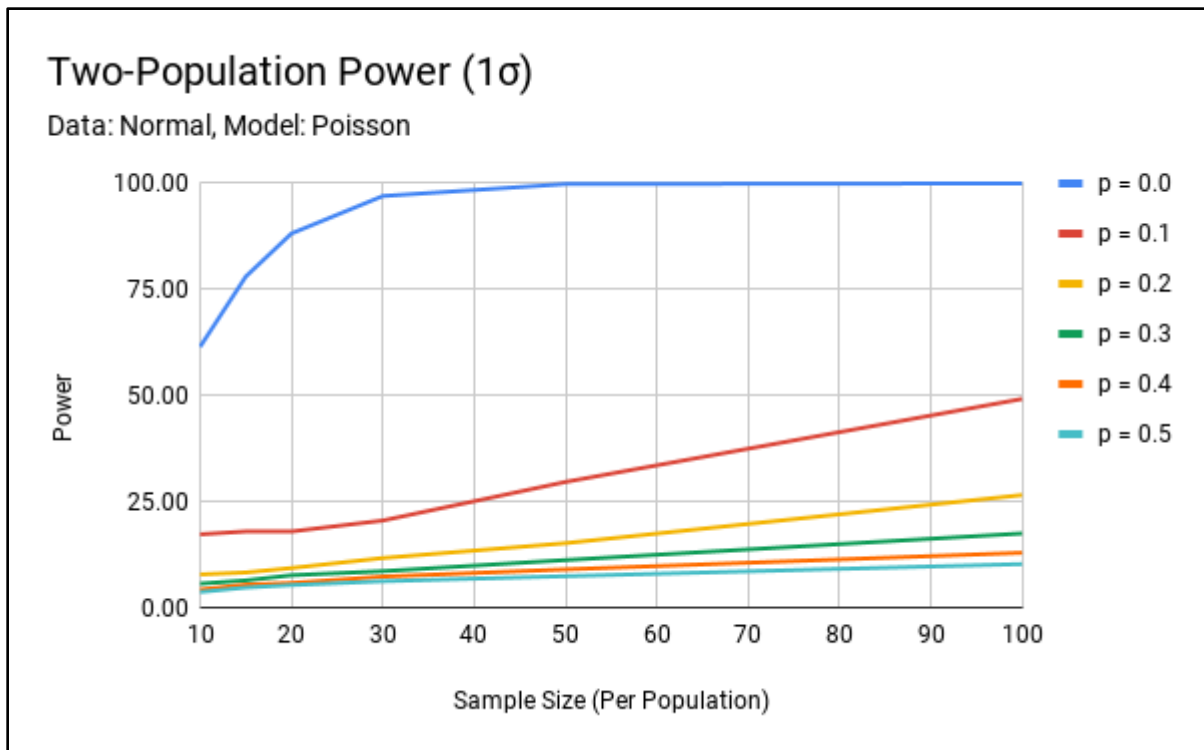


Figure B2. Normal Sample – Poisson Model Two-Population Power (1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 22$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 1.13 | 0.76 | 0.67 | 0.51 | 0.28 | 0.24 |
| n = 15 | 3.89 | 2.74 | 1.87 | 1.21 | 0.90 | 0.59 |
| n = 20 | 9.48 | 6.35 | 4.89 | 3.21 | 1.93 | 1.15 |
| n = 30 | 26.94 | 20.66 | 14.73 | 10.74 | 6.80 | 3.72 |
| n = 50 | 68.45 | 59.40 | 48.78 | 37.4 | 27.06 | 16.37 |
| n = 100 | 99.53 | 98.46 | 96.66 | 92.01 | 83.01 | 68.05 |

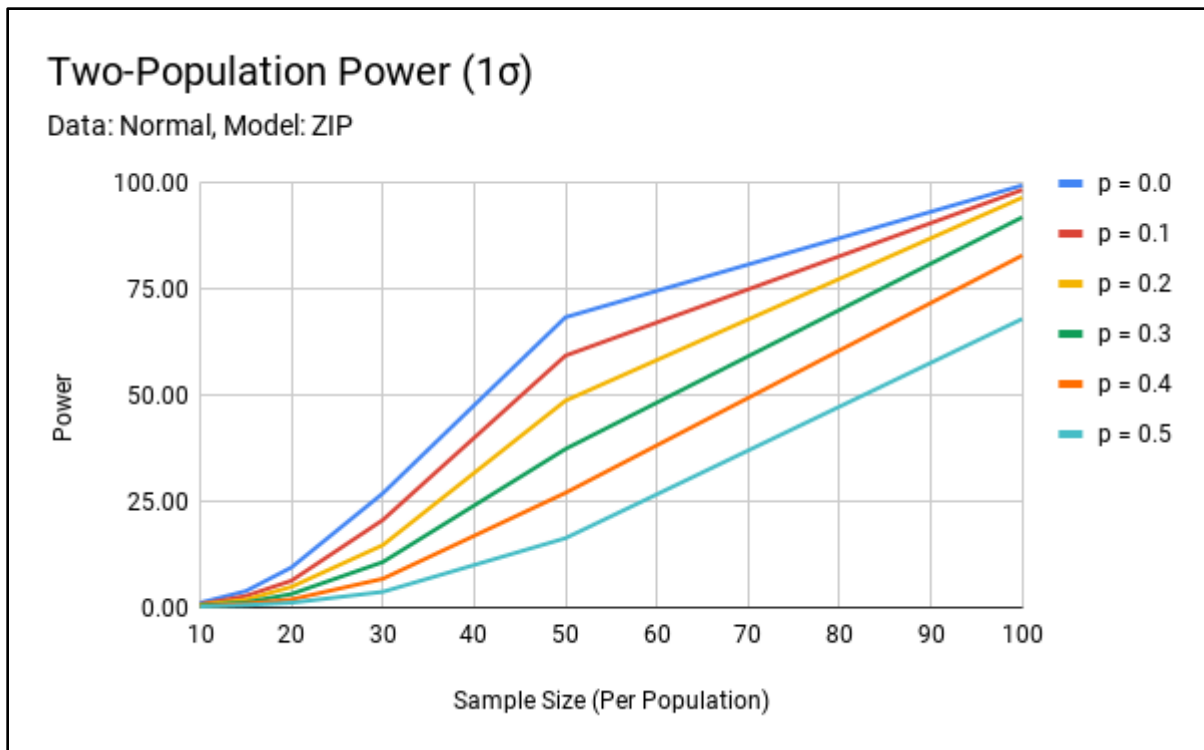


Figure B3. Normal Sample – ZIP Model Two-Population Power (1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 24.5$

$\lambda_2 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 52.68 | 23.92 | 13.99 | 9.78 | 8.00 | 6.55 |
| n = 15 | 71.15 | 29.28 | 17.10 | 12.36 | 8.95 | 7.47 |
| n = 20 | 83.34 | 35.96 | 21.10 | 14.12 | 11.09 | 8.83 |
| n = 30 | 95.51 | 49.07 | 28.87 | 19.53 | 14.16 | 10.91 |
| n = 50 | 99.73 | 70.23 | 43.45 | 29.21 | 21.03 | 15.74 |
| n = 100 | 100.00 | 93.49 | 72.59 | 52.58 | 37.51 | 26.77 |

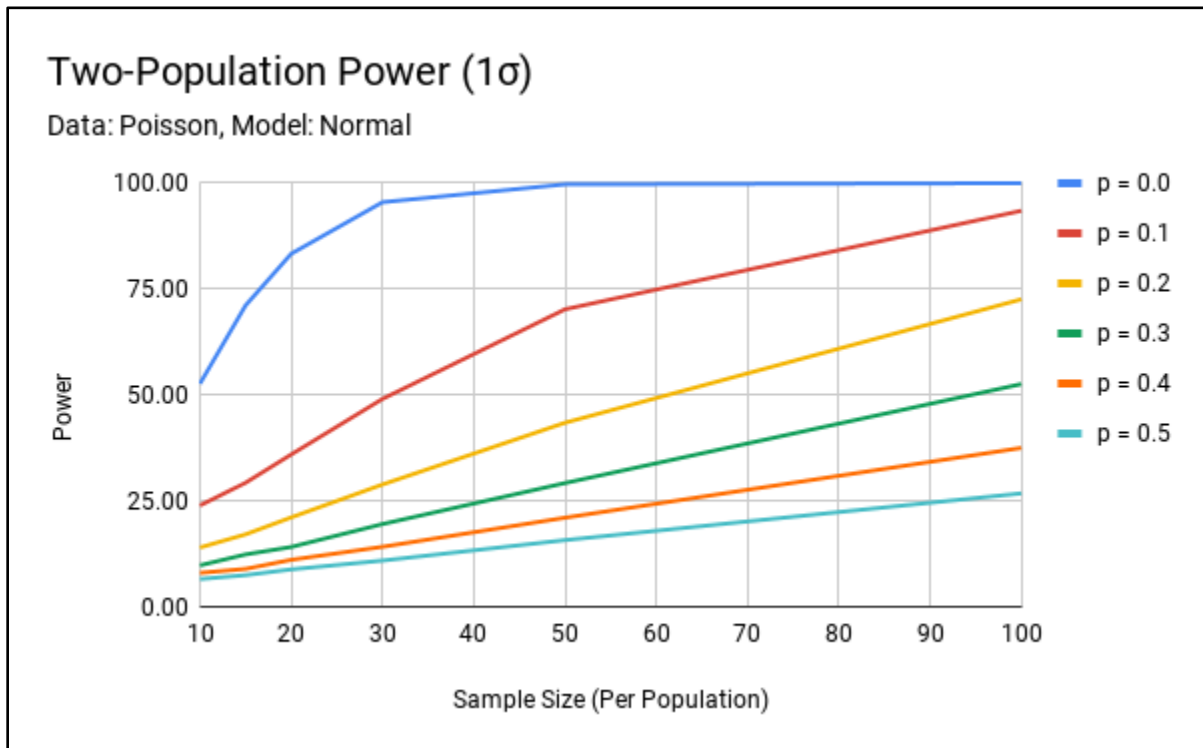


Figure B4. Poisson Sample – Normal Model Two-Population Power (1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 24.5$

$\lambda_2 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 57.86 | 25.62 | 14.54 | 9.57 | 7.13 | 5.39 |
| n = 15 | 74.09 | 31.18 | 18.01 | 13.07 | 9.30 | 7.28 |
| n = 20 | 85.43 | 38.40 | 22.38 | 14.91 | 11.22 | 8.48 |
| n = 30 | 95.53 | 51.06 | 29.80 | 19.79 | 15.04 | 11.12 |
| n = 50 | 99.78 | 70.74 | 44.62 | 30.11 | 21.46 | 15.80 |
| n = 100 | 100.00 | 93.48 | 73.04 | 52.48 | 38.19 | 27.85 |

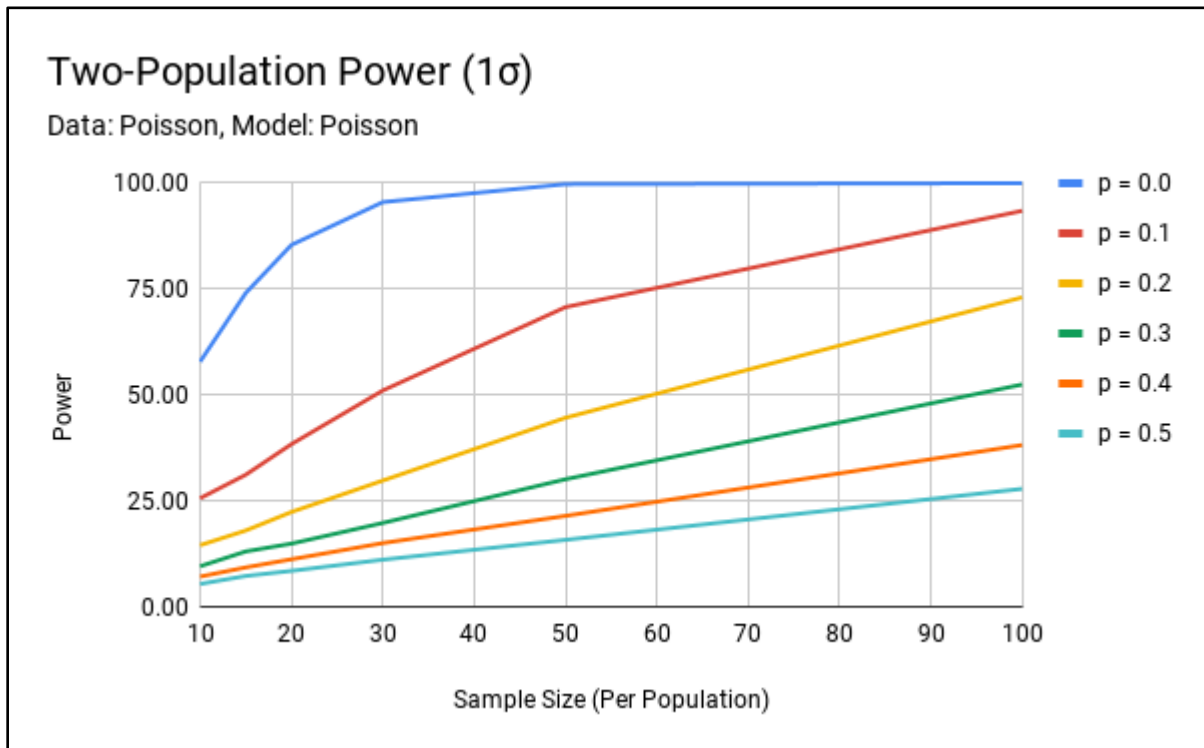


Figure B5. Poisson Sample – Poisson Model Two-Population Power (1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 24.5$

$\lambda_2 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 56.44 | 51.64 | 47.68 | 42.07 | 37.01 | 30.35 |
| n = 15 | 74.17 | 69.18 | 63.77 | 58.49 | 50.10 | 44.19 |
| n = 20 | 85.75 | 81.34 | 77.14 | 70.13 | 63.94 | 55.38 |
| n = 30 | 95.73 | 93.80 | 90.82 | 86.37 | 80.16 | 73.01 |
| n = 50 | 99.80 | 99.45 | 98.91 | 97.52 | 95.76 | 91.36 |
| n = 100 | 100.00 | 100.00 | 100.00 | 99.99 | 99.93 | 99.77 |

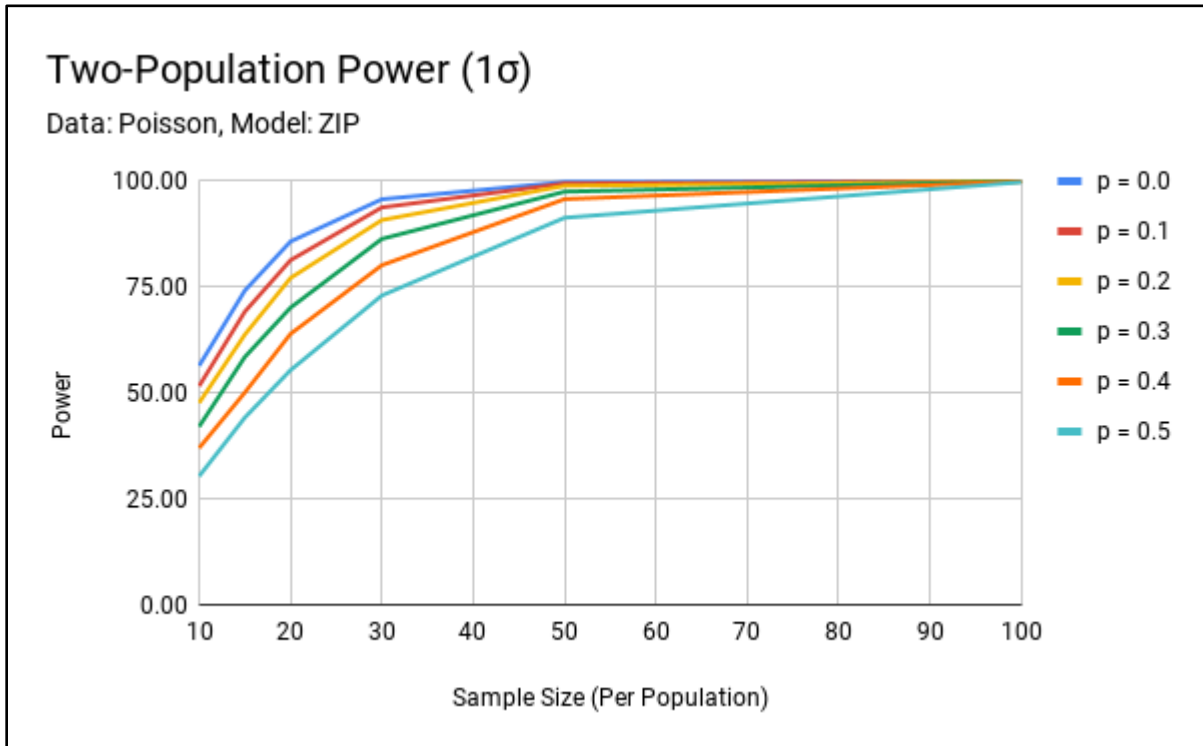


Figure B6. Poisson Sample – ZIP Model Two-Population Power (1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 26.3$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 51.43 | 27.08 | 16.65 | 12.26 | 9.41 | 7.42 |
| n = 15 | 68.56 | 36.62 | 23.66 | 16.30 | 11.91 | 9.35 |
| n = 20 | 81.93 | 45.32 | 28.57 | 20.51 | 14.47 | 11.90 |
| n = 30 | 94.24 | 61.21 | 41.28 | 28.19 | 21.10 | 14.73 |
| n = 50 | 99.57 | 83.29 | 60.09 | 43.10 | 30.85 | 22.50 |
| n = 100 | 100.00 | 98.55 | 88.59 | 72.17 | 56.58 | 41.04 |

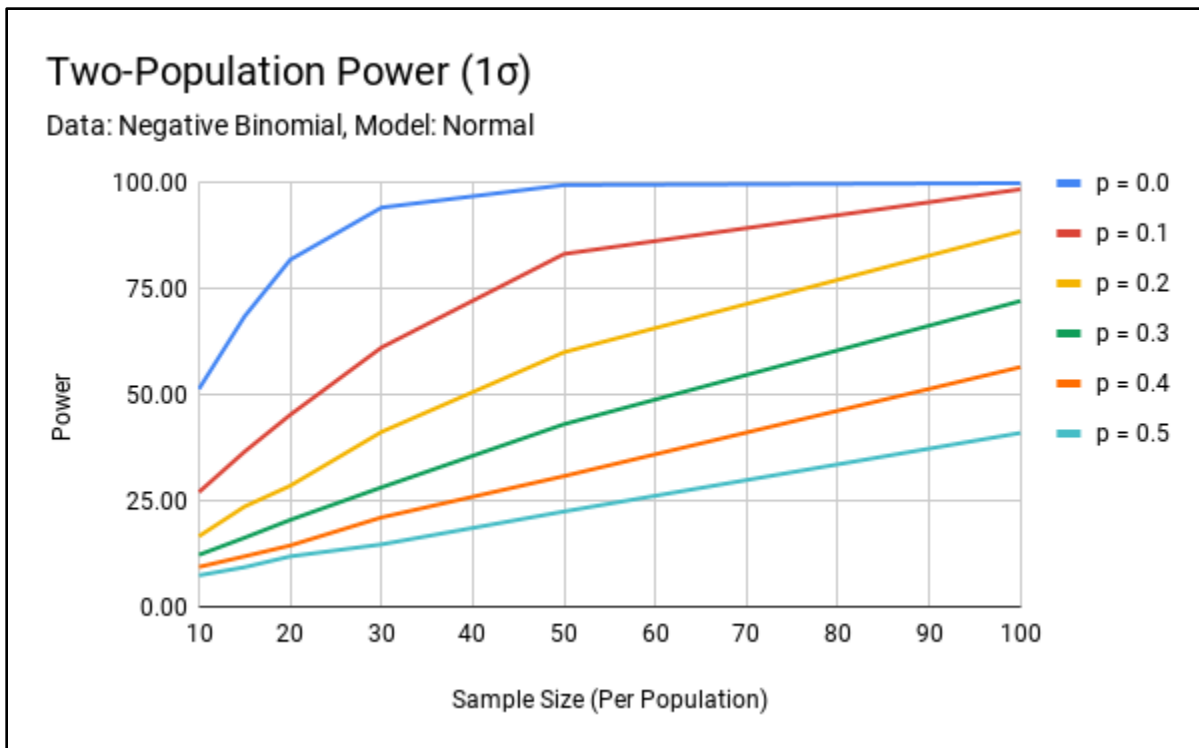


Figure B7. Negative Binomial Sample – Normal Model Two-Population Power (1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 26.3$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 55.08 | 30.30 | 18.60 | 12.49 | 9.47 | 6.54 |
| n = 15 | 71.59 | 38.88 | 24.15 | 17.08 | 12.32 | 9.78 |
| n = 20 | 82.98 | 48.04 | 30.88 | 20.98 | 15.71 | 11.88 |
| n = 30 | 94.27 | 64.15 | 43.66 | 28.86 | 21.34 | 15.39 |
| n = 50 | 99.60 | 84.22 | 61.54 | 44.60 | 31.67 | 23.15 |
| n = 100 | 100.00 | 98.57 | 89.01 | 74.38 | 56.53 | 42.03 |

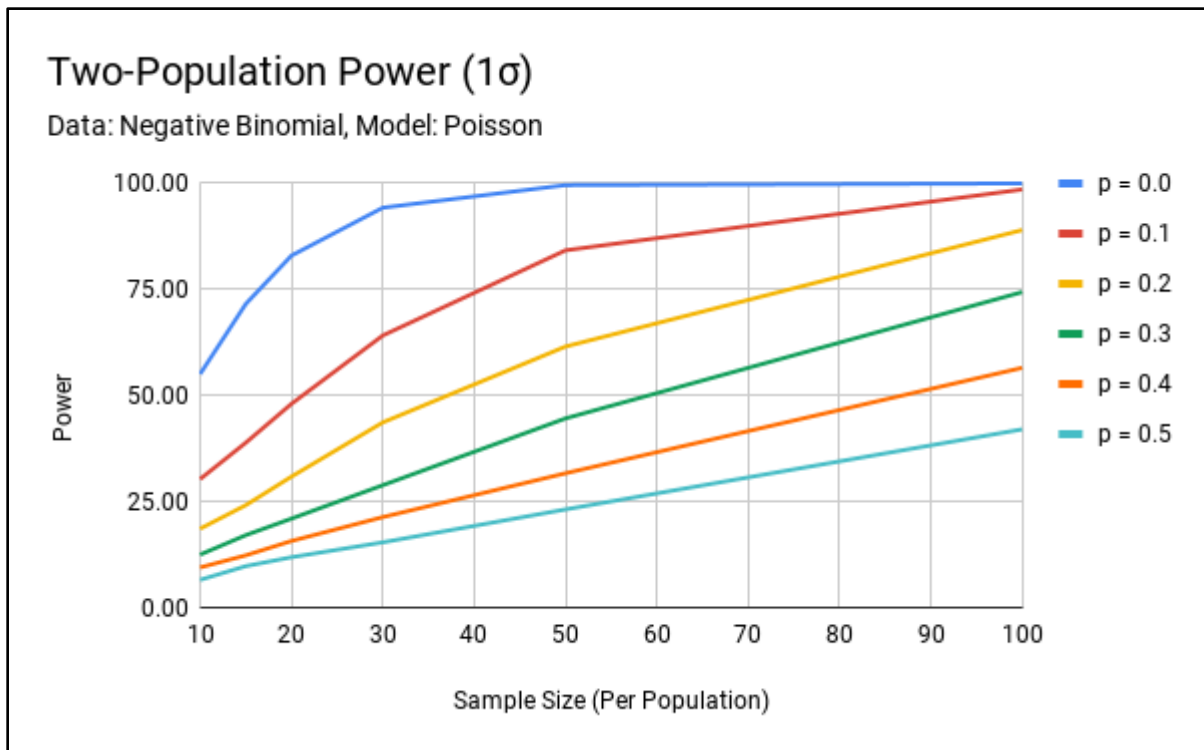


Figure B8. Negative Binomial Sample – Poisson Model Two-Population Power (1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 26.3$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 75.59 | 71.49 | 67.15 | 62.17 | 56.82 | 51.70 |
| n = 15 | 87.91 | 84.11 | 80.56 | 75.84 | 70.24 | 64.39 |
| n = 20 | 93.64 | 92.20 | 88.79 | 84.88 | 80.52 | 73.88 |
| n = 30 | 98.43 | 97.82 | 96.53 | 94.85 | 90.69 | 86.95 |
| n = 50 | 99.93 | 99.86 | 99.74 | 99.23 | 98.34 | 96.81 |
| n = 100 | 100.00 | 100.00 | 100.00 | 100.00 | 99.99 | 99.90 |

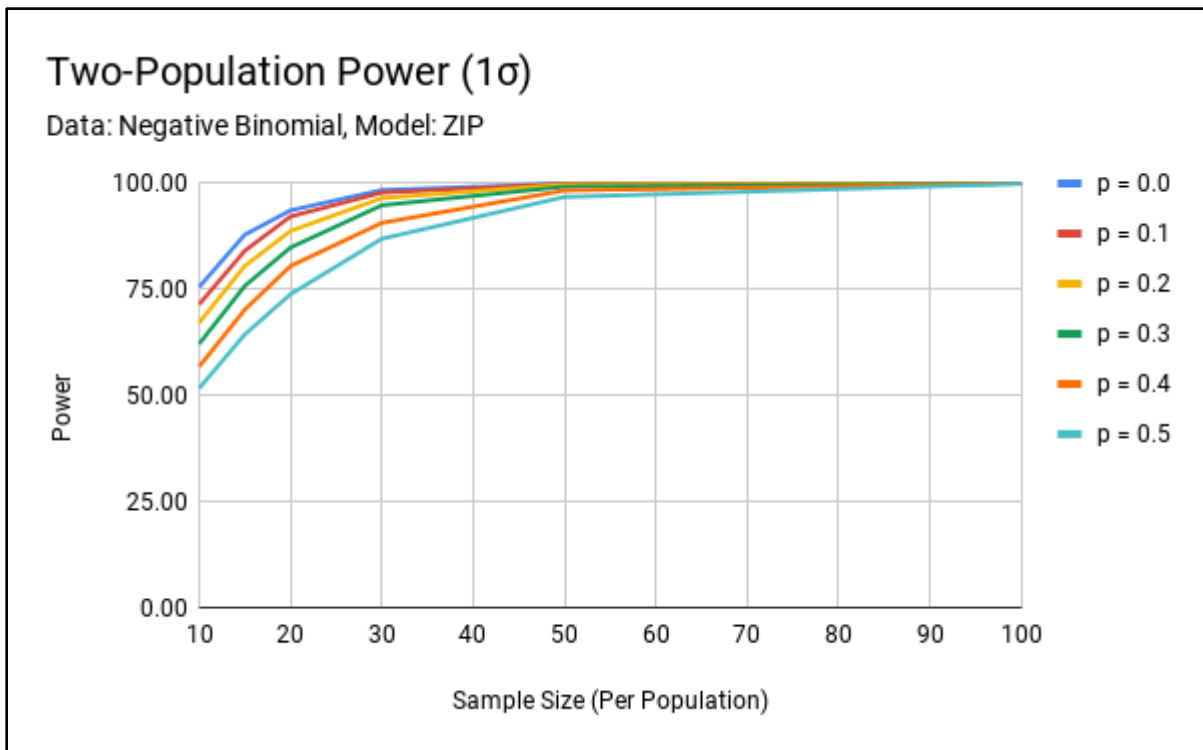


Figure B9. Negative Binomial Sample – ZIP Model Two-Population Power (1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 24$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 98.79 | 35.44 | 18.81 | 13.71 | 12.03 | 6.82 |
| n = 15 | 99.98 | 32.42 | 21.19 | 14.43 | 11.83 | 7.18 |
| n = 20 | 100.00 | 41.42 | 23.40 | 16.40 | 12.75 | 8.32 |
| n = 30 | 100.00 | 55.82 | 30.69 | 19.86 | 14.62 | 12.11 |
| n = 50 | 100.00 | 74.31 | 43.86 | 28.85 | 20.72 | 15.21 |
| n = 100 | 100.00 | 94.72 | 70.68 | 49.25 | 34.60 | 24.58 |

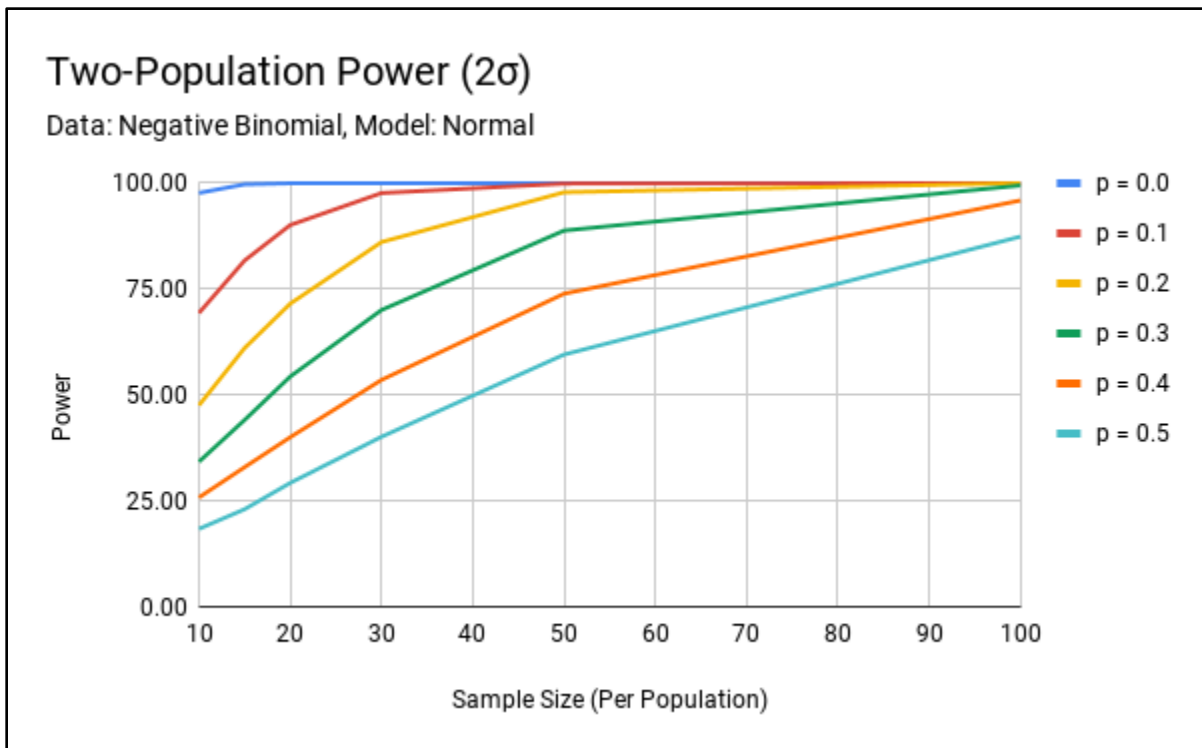


Figure B10. Normal Sample – Normal Model Two-Population Power (2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 24$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 99.27 | 35.34 | 15.09 | 9.11 | 7.19 | 5.16 |
| n = 15 | 99.93 | 34.98 | 17.82 | 11.87 | 8.83 | 6.43 |
| n = 20 | 100.00 | 42.66 | 21.27 | 14.22 | 10.50 | 7.89 |
| n = 30 | 100.00 | 55.01 | 29.37 | 18.38 | 13.03 | 10.52 |
| n = 50 | 100.00 | 74.21 | 43.18 | 28.09 | 19.76 | 14.29 |
| n = 100 | 100.00 | 94.79 | 70.67 | 49.09 | 34.29 | 24.15 |

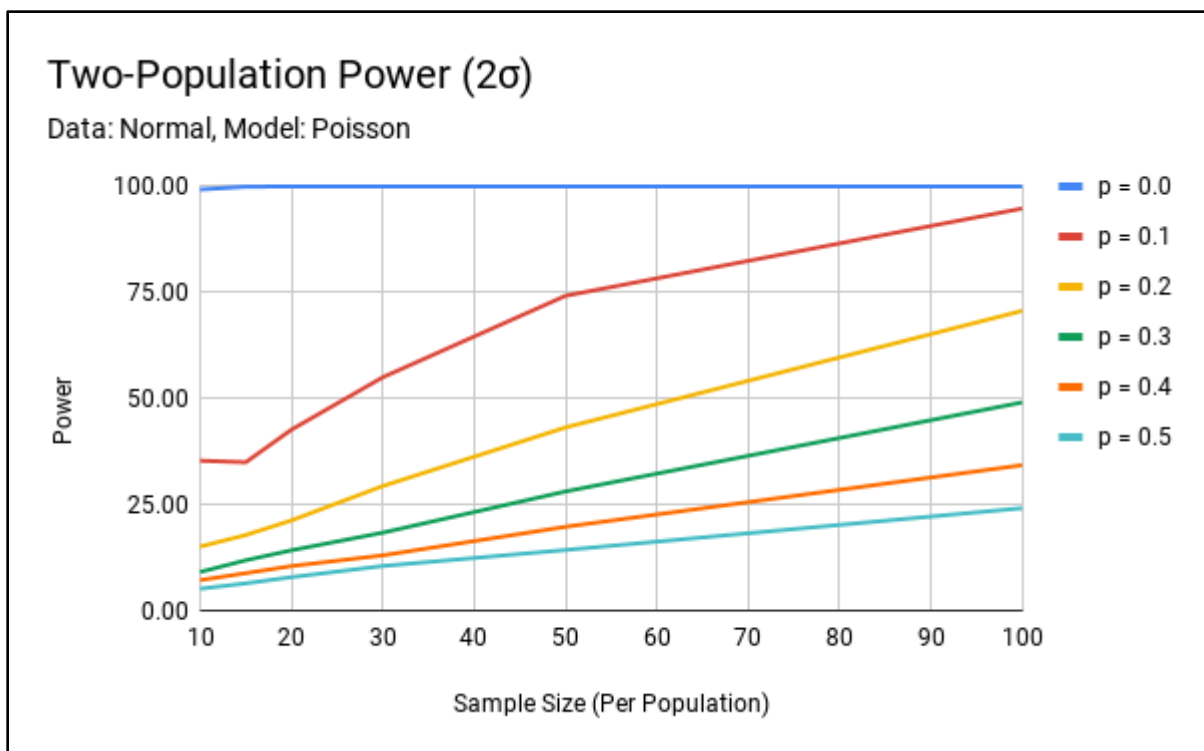


Figure B11. Normal Sample – Poisson Model Two-Population Power (2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\mu_1 = 24$

$\mu_2 = 20$

$\sigma_1 = \sigma_2 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 45.33 | 35.41 | 26.11 | 19.06 | 11.99 | 7.41 |
| n = 15 | 80.76 | 71.44 | 60.20 | 47.60 | 34.44 | 22.70 |
| n = 20 | 95.70 | 91.19 | 84.05 | 72.43 | 59.48 | 42.34 |
| n = 30 | 99.92 | 99.61 | 98.79 | 96.54 | 90.04 | 76.88 |
| n = 50 | 100.00 | 100.00 | 100.00 | 99.99 | 99.81 | 98.74 |
| n = 100 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

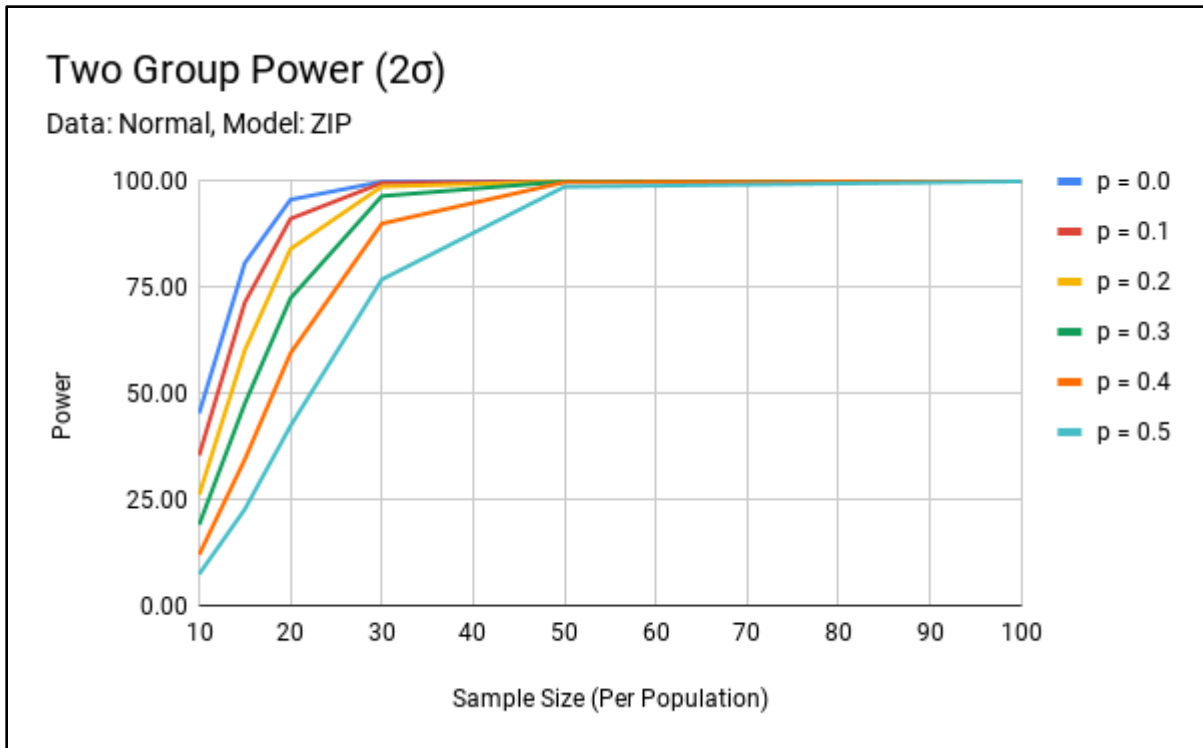


Figure B12. Normal Sample – ZIP Model Two-Population Power (2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 29$

$\lambda_2 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 98.51 | 58.13 | 37.85 | 25.97 | 20.09 | 15.04 |
| n = 15 | 99.93 | 70.95 | 47.10 | 32.60 | 23.72 | 17.56 |
| n = 20 | 100.00 | 80.41 | 57.32 | 40.36 | 28.23 | 21.14 |
| n = 30 | 100.00 | 92.40 | 72.41 | 53.75 | 38.21 | 27.43 |
| n = 50 | 100.00 | 99.01 | 89.98 | 74.07 | 56.17 | 41.23 |
| n = 100 | 100.00 | 100.00 | 99.35 | 94.84 | 84.23 | 69.99 |

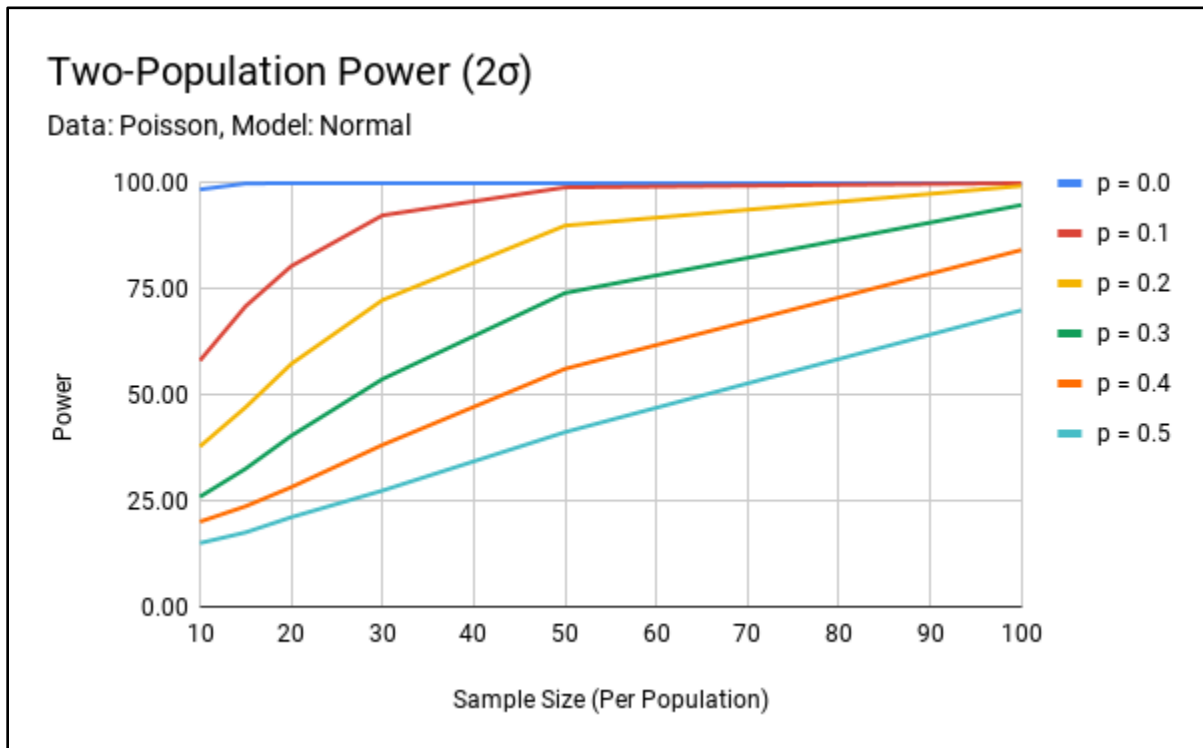


Figure B13. Poisson Sample – Normal Model Two-Population Power (2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 29$

$\lambda_2 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 98.06 | 55.32 | 33.98 | 21.69 | 15.40 | 10.48 |
| n = 15 | 99.90 | 69.95 | 44.91 | 30.17 | 21.00 | 14.85 |
| n = 20 | 100.00 | 80.07 | 56.29 | 38.99 | 26.35 | 19.21 |
| n = 30 | 100.00 | 92.40 | 72.29 | 53.22 | 37.41 | 26.26 |
| n = 50 | 100.00 | 99.03 | 90.19 | 74.34 | 56.19 | 41.08 |
| n = 100 | 100.00 | 100.00 | 99.38 | 95.03 | 84.53 | 70.52 |

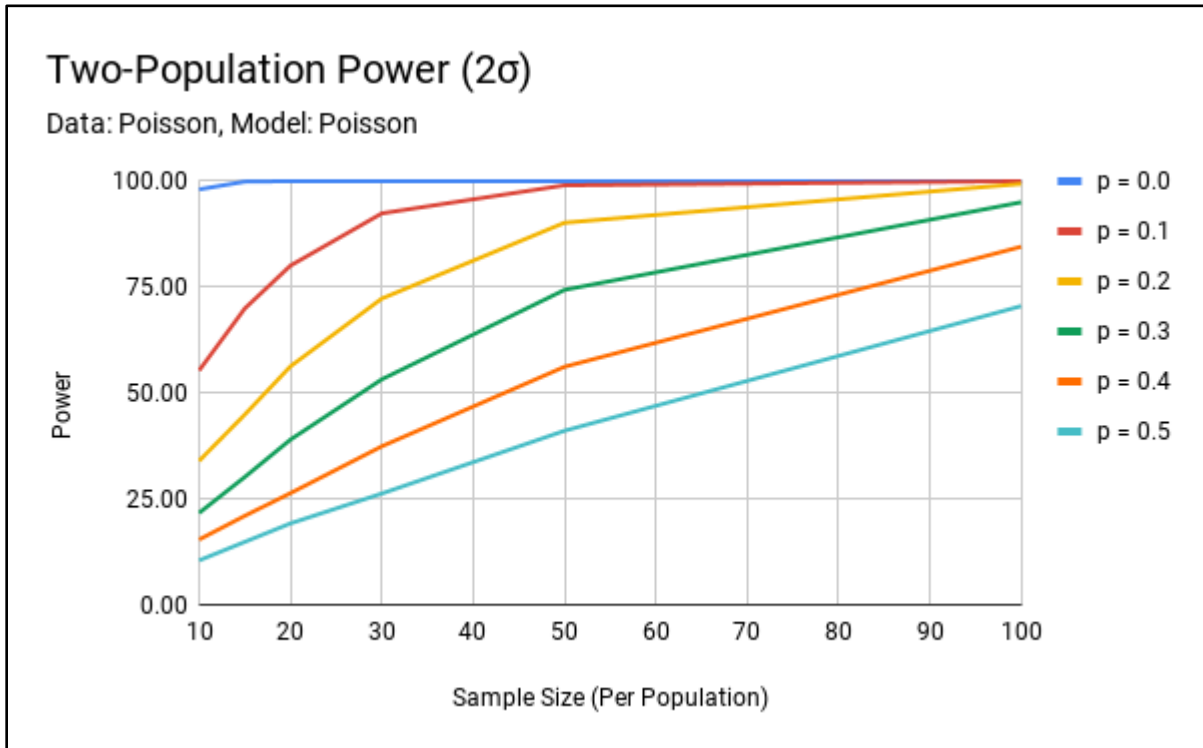


Figure B14. Poisson Sample – Poisson Model Two-Population Power (2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$\lambda_1 = 29$

$\lambda_2 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 98.38 | 97.12 | 94.48 | 91.48 | 85.47 | 77.51 |
| n = 15 | 99.91 | 99.57 | 99.25 | 98.10 | 96.14 | 92.08 |
| n = 20 | 100.00 | 99.97 | 99.92 | 99.76 | 99.04 | 97.44 |
| n = 30 | 100.00 | 100.00 | 99.99 | 99.99 | 99.93 | 99.82 |
| n = 50 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 100 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

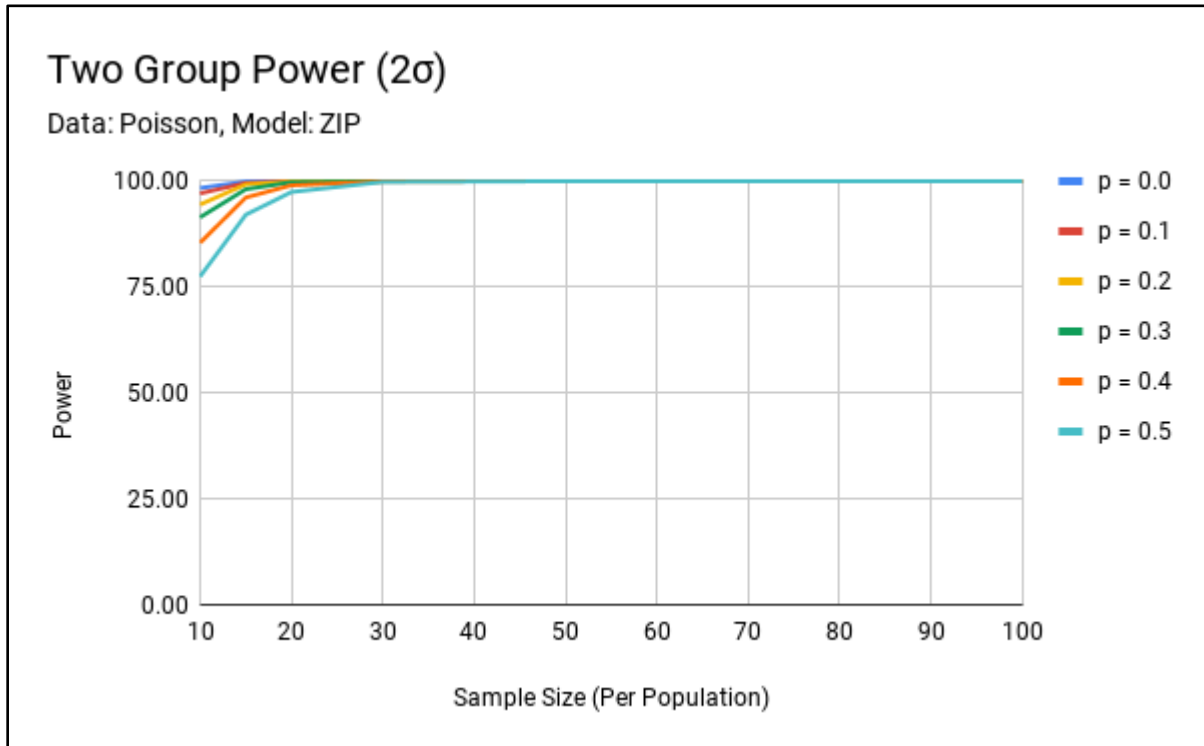


Figure B15. Poisson Sample – ZIP Model Two-Population Power (2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 32.6$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 97.69 | 69.36 | 47.57 | 34.25 | 25.81 | 18.39 |
| n = 15 | 99.73 | 81.74 | 61.08 | 44.10 | 32.96 | 23.03 |
| n = 20 | 99.98 | 90.12 | 71.55 | 54.35 | 40.03 | 29.24 |
| n = 30 | 99.99 | 97.67 | 86.10 | 70.06 | 53.56 | 40.13 |
| n = 50 | 100.00 | 99.90 | 97.86 | 88.81 | 73.91 | 59.54 |
| n = 100 | 100.00 | 100.00 | 100.00 | 99.47 | 95.93 | 87.39 |

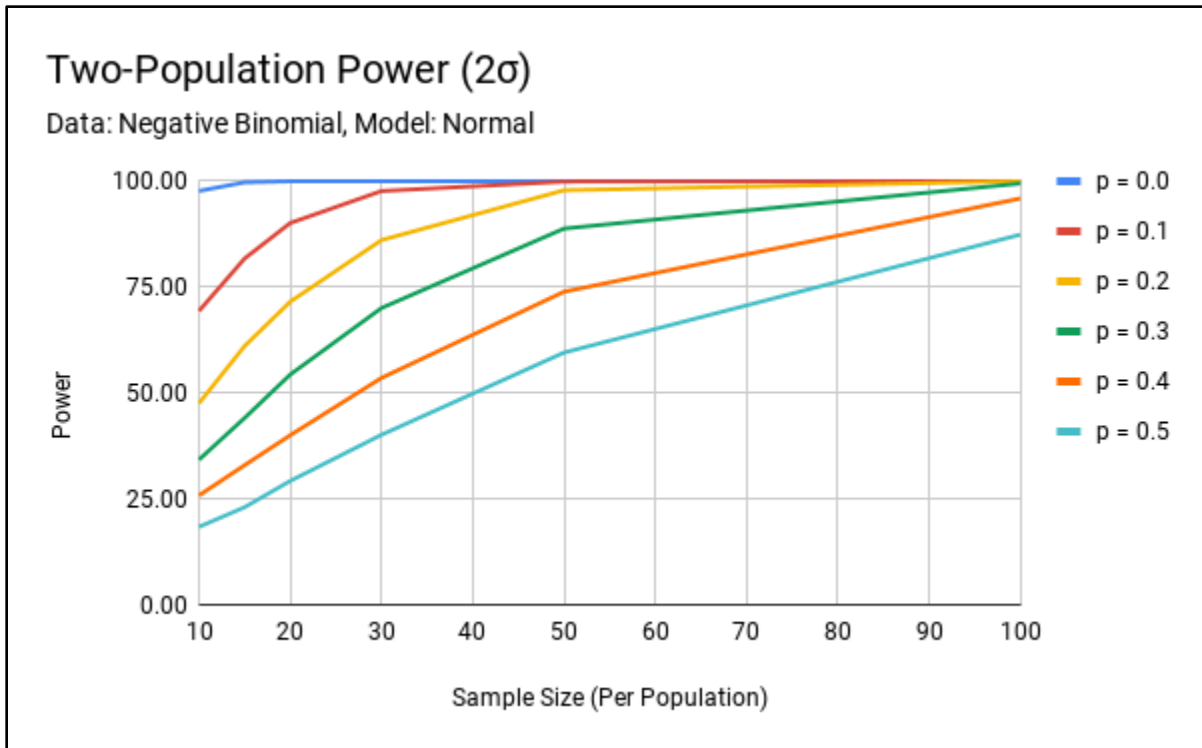


Figure B16. Negative Binomial Sample – Normal Model Two-Population Power (2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 32.6$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 97.12 | 66.99 | 44.49 | 30.66 | 21.28 | 13.54 |
| n = 15 | 99.63 | 81.10 | 59.91 | 42.27 | 30.52 | 20.36 |
| n = 20 | 99.97 | 90.06 | 71.06 | 53.21 | 38.65 | 27.80 |
| n = 30 | 99.99 | 97.70 | 86.26 | 70.10 | 53.29 | 39.90 |
| n = 50 | 100.00 | 99.92 | 97.91 | 89.11 | 74.39 | 59.99 |
| n = 100 | 100.00 | 100.00 | 100.00 | 99.49 | 96.14 | 87.82 |

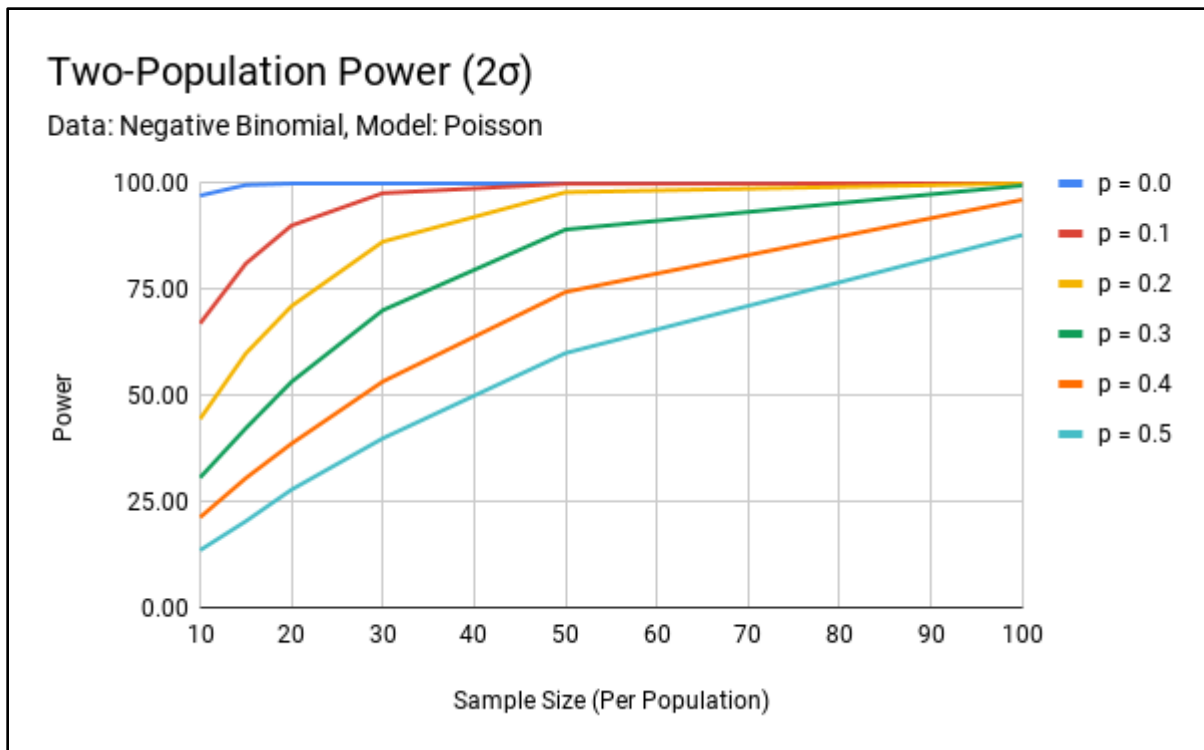


Figure B17. Negative Binomial Sample – Poisson Model Two-Population Power (2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2$

$p_1 = p_2$

$k(nb)_1 = 32.6$

$k(nb)_2 = 20$

$p(nb)_1 = p(nb)_2 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 10 | 99.53 | 98.85 | 97.65 | 96.03 | 92.97 | 88.15 |
| n = 15 | 99.94 | 99.91 | 99.78 | 99.27 | 98.41 | 96.90 |
| n = 20 | 100.00 | 100.00 | 99.97 | 99.97 | 99.70 | 98.90 |
| n = 30 | 100.00 | 100.00 | 100.00 | 99.99 | 99.97 | 99.91 |
| n = 50 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 100 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

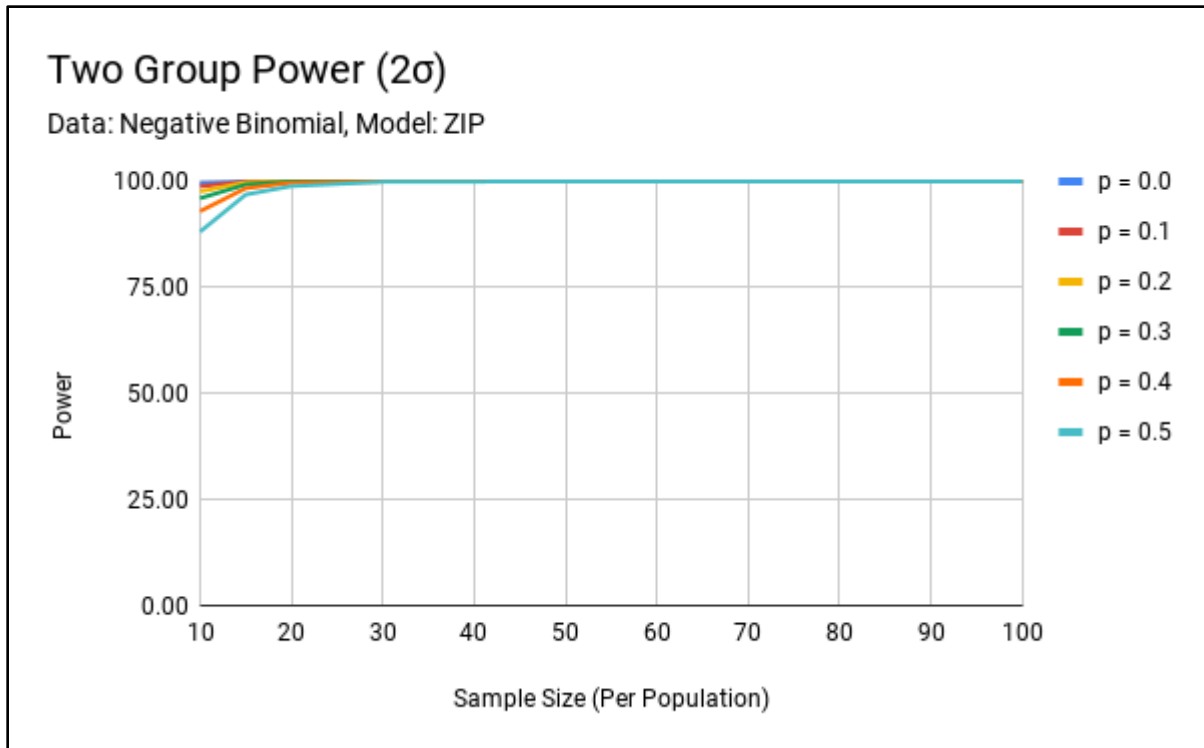


Figure B18. Negative Binomial Sample – ZIP Model Two-Population Power (2σ).

Eight-Population 1 vs 7 Scenario

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 22$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.72 | 5.21 | 5.33 | 5.70 | 5.75 | 5.93 |
| n = 5 | 22.33 | 5.54 | 5.26 | 5.46 | 5.91 | 5.70 |
| n = 10 | 49.57 | 7.05 | 6.13 | 5.86 | 5.54 | 5.20 |

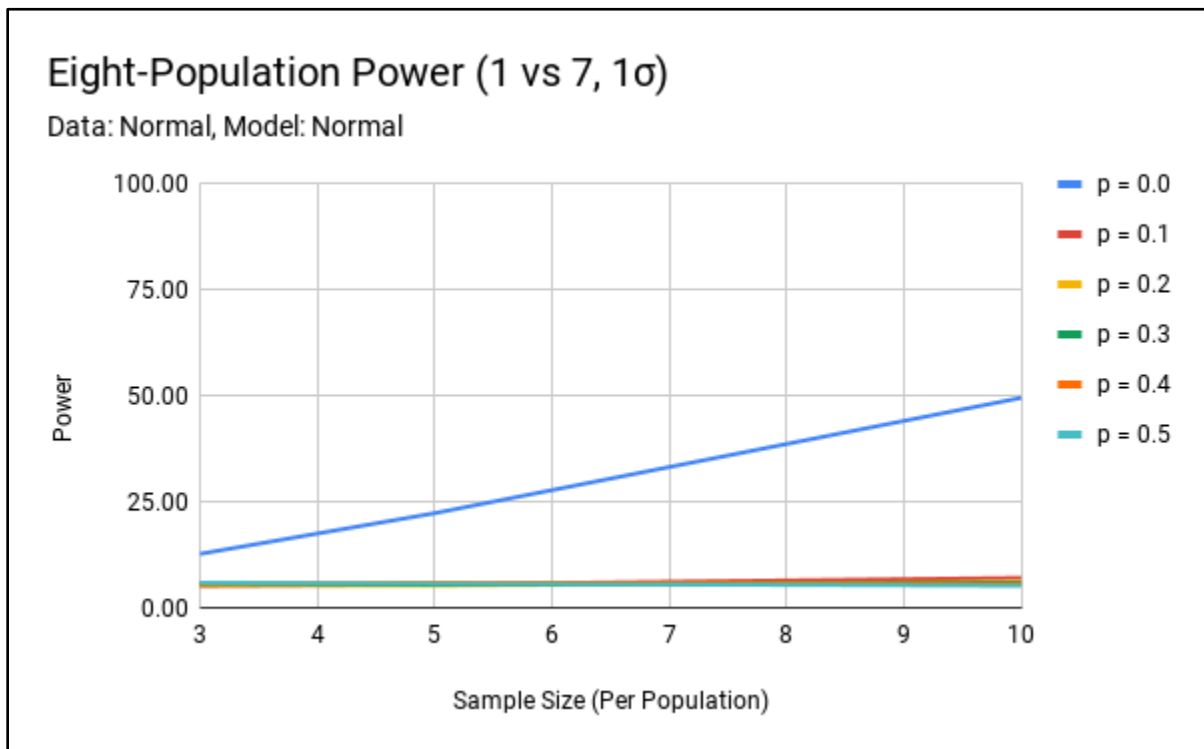


Figure B19. Normal Sample – Normal Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 22$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 24.87 | 3.83 | 6.05 | 12.51 | 14.22 | 11.63 |
| n = 5 | 30.11 | 3.68 | 3.09 | 3.91 | 7.21 | 10.71 |
| n = 10 | 54.19 | 5.62 | 4.65 | 4.92 | 4.82 | 5.01 |

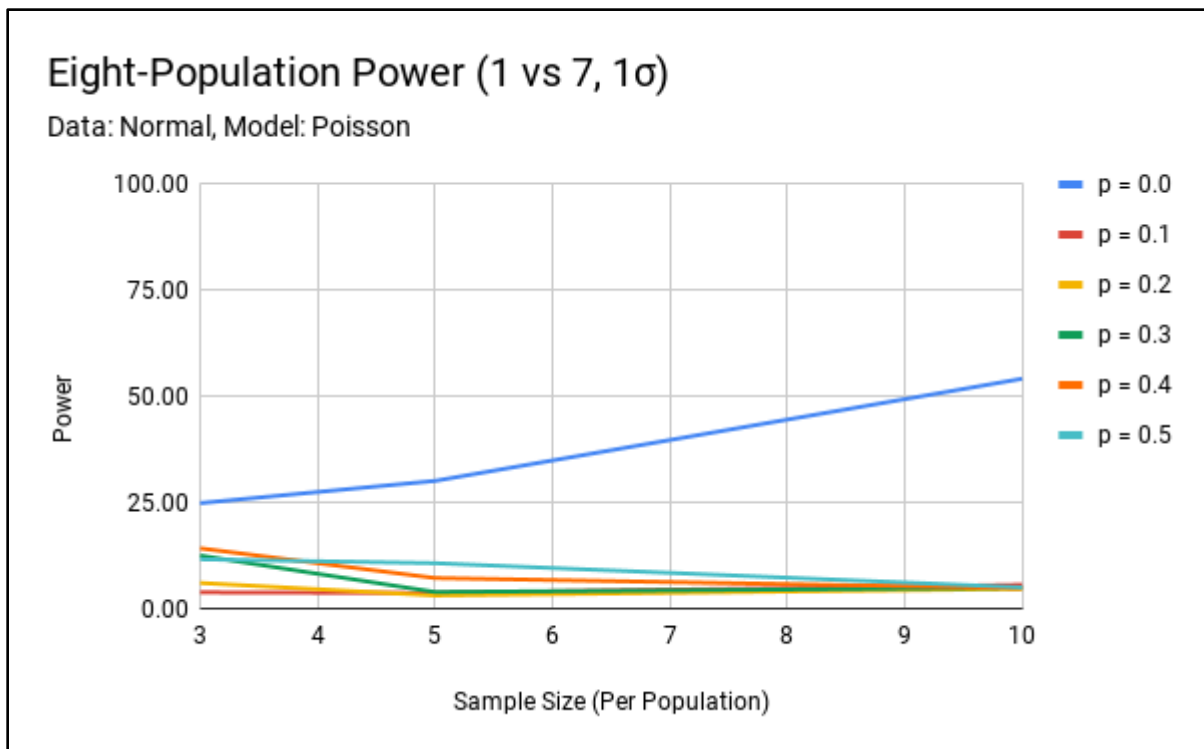


Figure B20. Normal Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 22$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 0.00 | 0.19 | 0.49 | 0.80 | 1.17 | 1.47 |
| n = 5 | 0.00 | 0.01 | 0.23 | 0.55 | 0.51 | 0.73 |
| n = 10 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.44 |

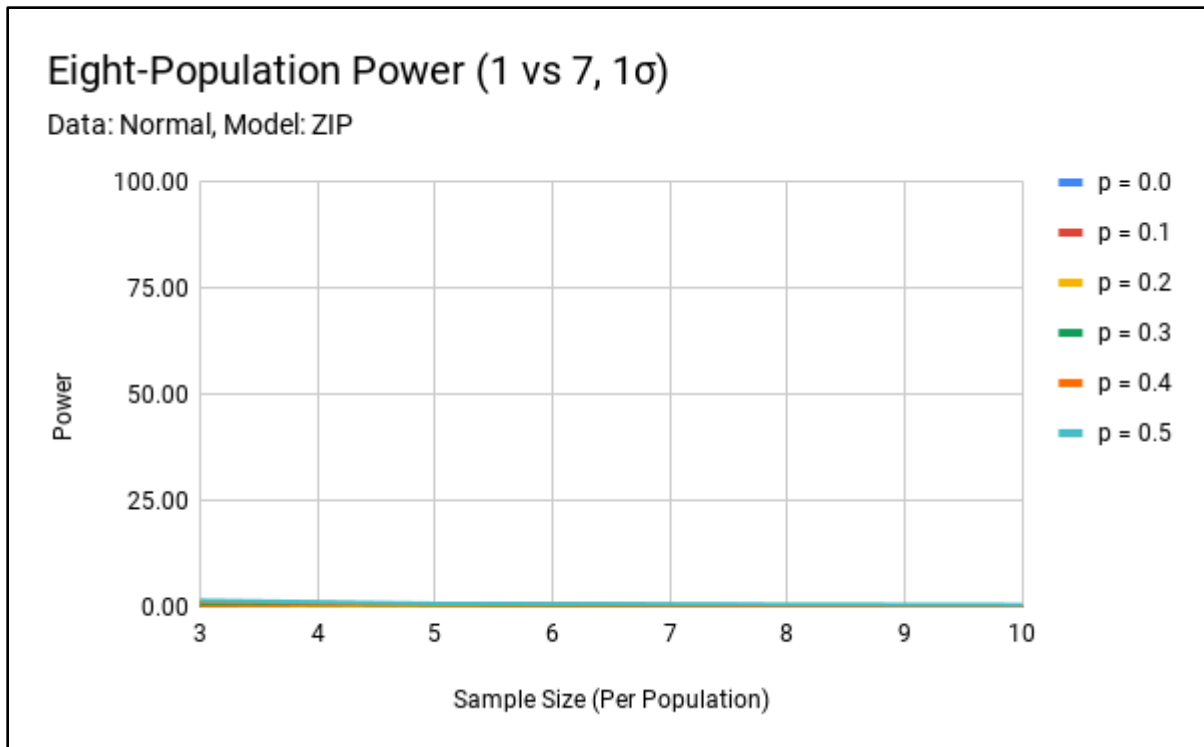


Figure B21. Normal Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 24.5$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 13.62 | 7.07 | 6.40 | 6.64 | 5.68 | 5.85 |
| n = 5 | 21.96 | 10.19 | 7.54 | 7.15 | 6.29 | 6.25 |
| n = 10 | 48.38 | 18.03 | 11.24 | 8.80 | 7.34 | 6.53 |

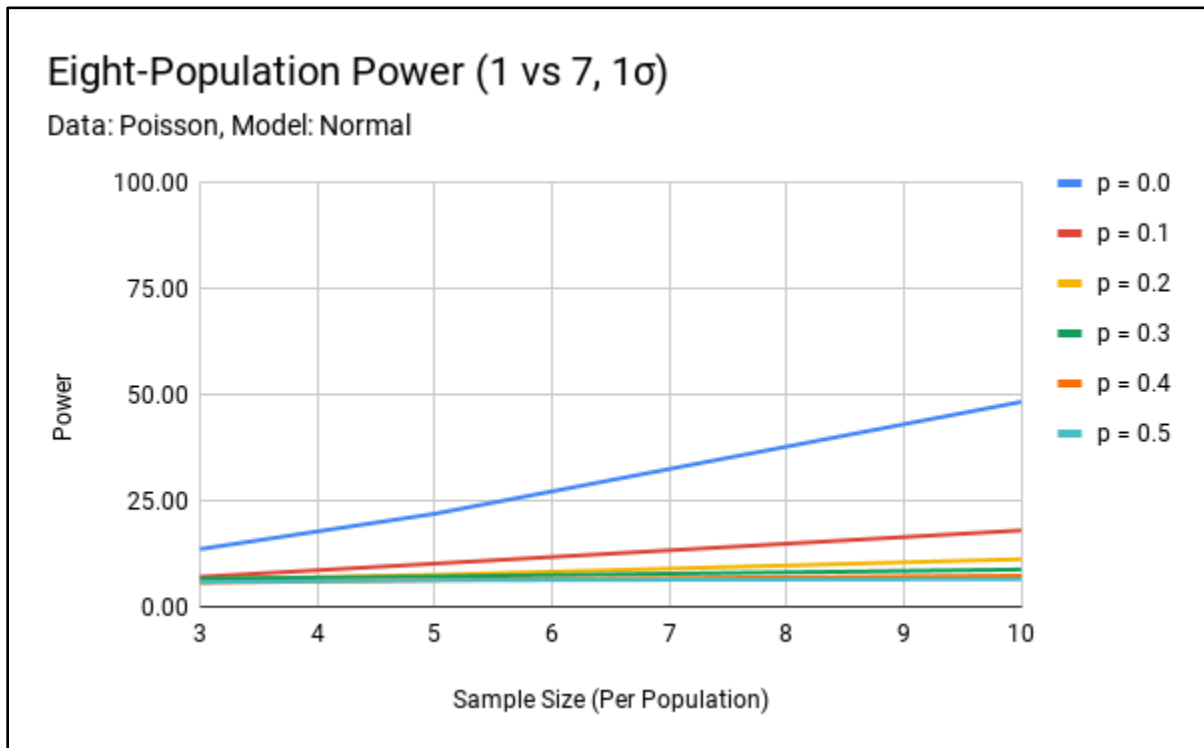


Figure B22. Poisson Sample – Normal Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 24.5$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 24.08 | 7.66 | 7.89 | 12.55 | 15.40 | 13.74 |
| n = 5 | 30.28 | 9.21 | 5.96 | 6.26 | 8.78 | 11.36 |
| n = 10 | 51.13 | 16.25 | 10.20 | 7.82 | 6.79 | 6.06 |

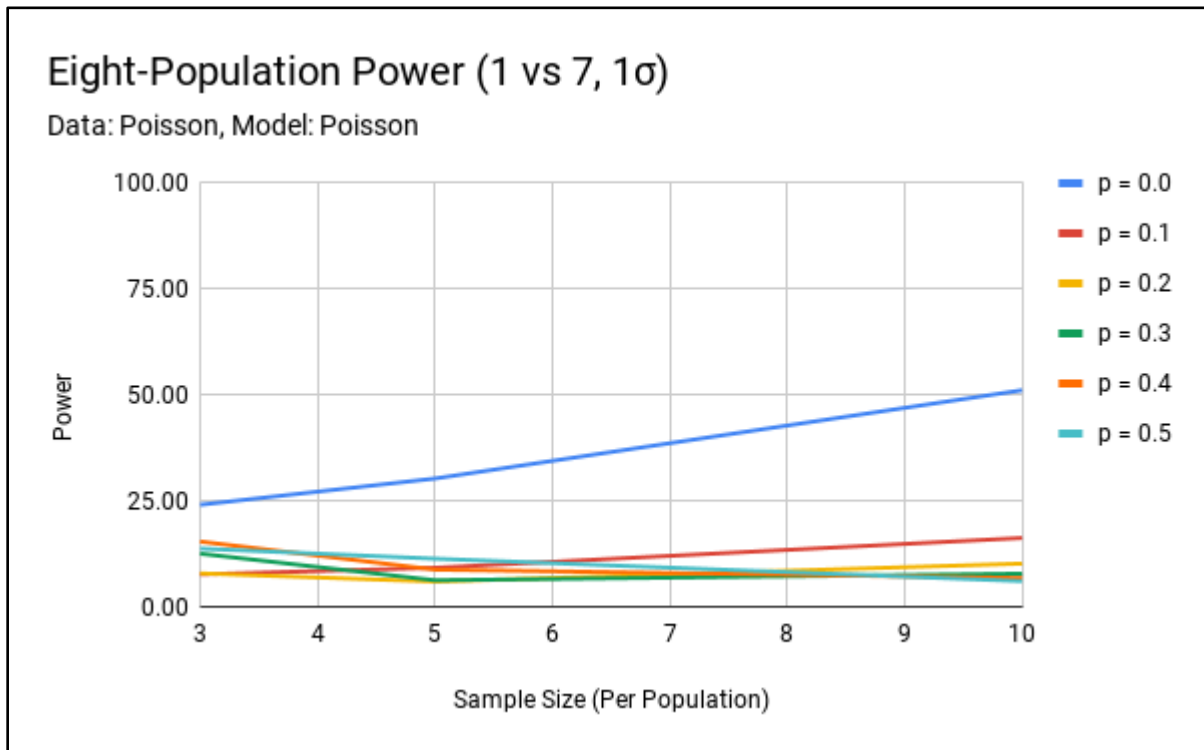


Figure B23. Poisson Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 24.5$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 16.42 | 14.93 | 17.49 | 19.91 | 22.86 | 24.54 |
| n = 5 | 26.24 | 23.40 | 20.81 | 20.21 | 20.63 | 22.02 |
| n = 10 | 50.69 | 46.51 | 41.26 | 36.14 | 30.55 | 26.82 |

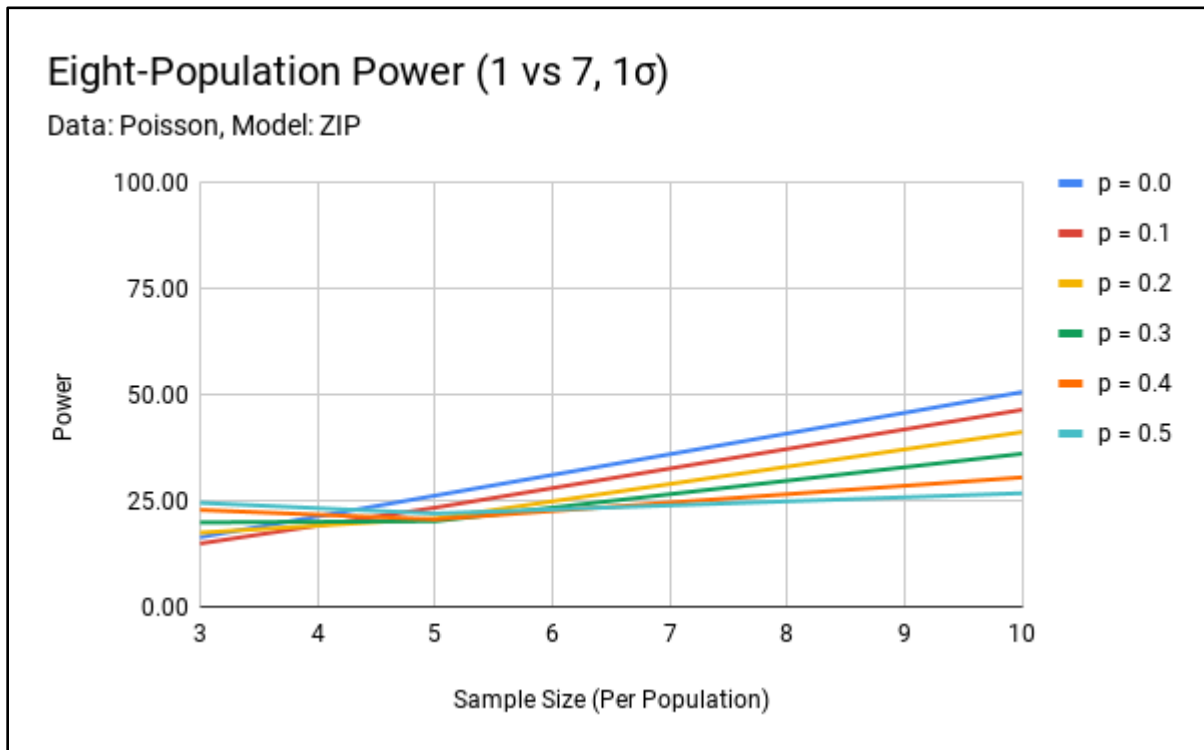


Figure B24. Poisson Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 26.3$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 12.70 | 8.38 | 7.26 | 6.69 | 6.41 | 6.36 |
| n = 5 | 21.75 | 12.98 | 9.65 | 7.68 | 7.53 | 6.19 |
| n = 10 | 46.77 | 23.66 | 14.90 | 11.12 | 9.37 | 8.30 |

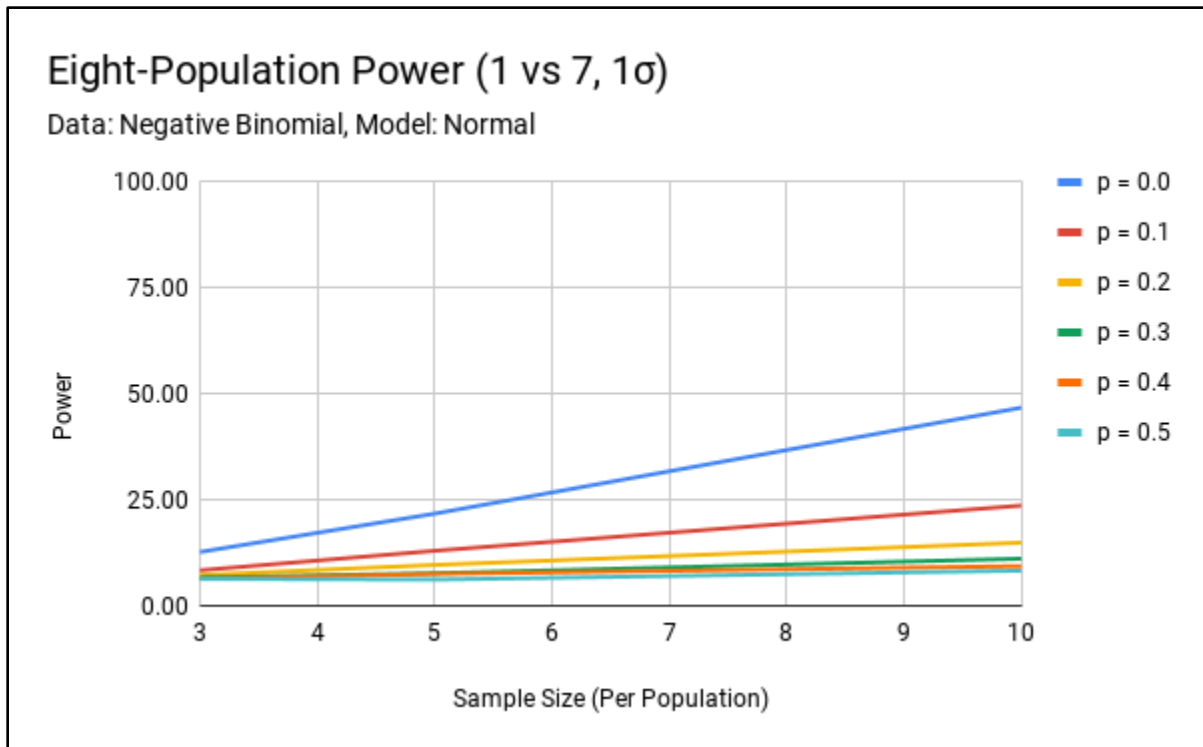


Figure B25. Negative Binomial Sample – Normal Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 26.3$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 23.61 | 11.96 | 10.80 | 14.12 | 16.05 | 15.25 |
| n = 5 | 30.06 | 13.63 | 9.40 | 8.45 | 10.30 | 12.40 |
| n = 10 | 50.43 | 23.17 | 14.77 | 11.03 | 9.36 | 8.25 |

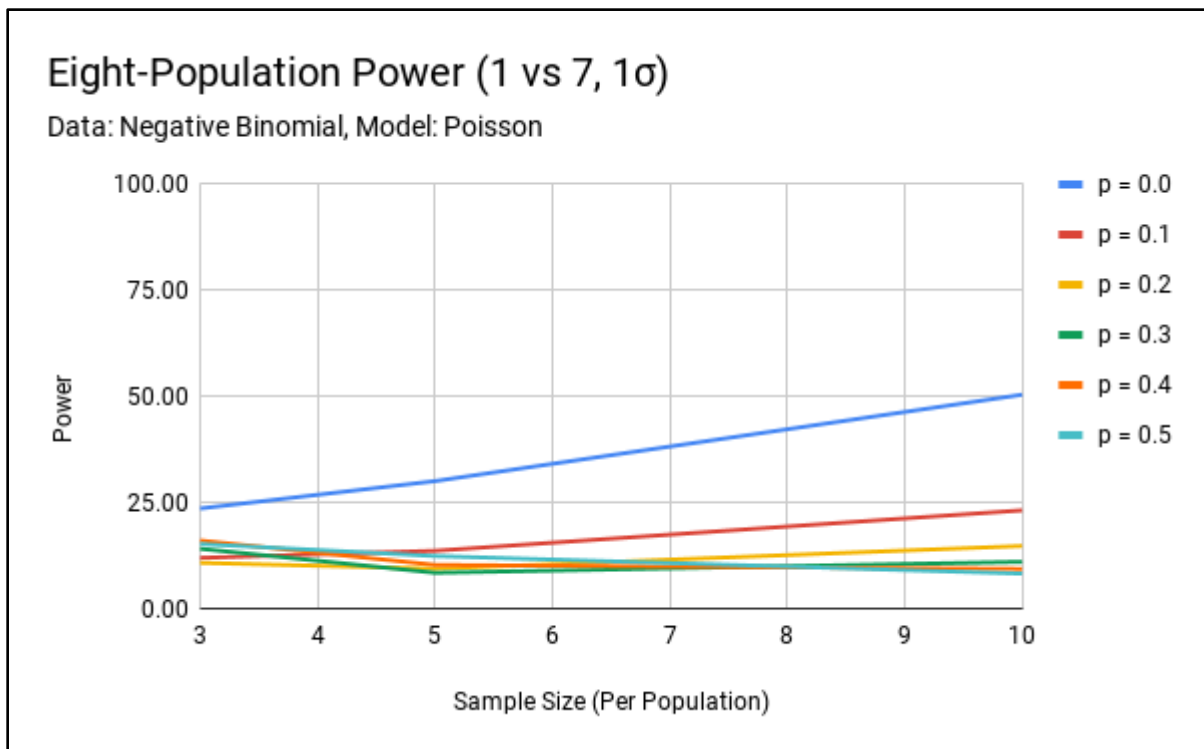


Figure B26. Negative Binomial Sample – Poisson Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 26.3$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 63.38 | 61.46 | 60.90 | 62.59 | 63.82 | 65.42 |
| n = 5 | 73.23 | 70.74 | 69.11 | 65.99 | 64.67 | 64.95 |
| n = 10 | 89.47 | 86.57 | 83.99 | 80.89 | 77.16 | 72.74 |

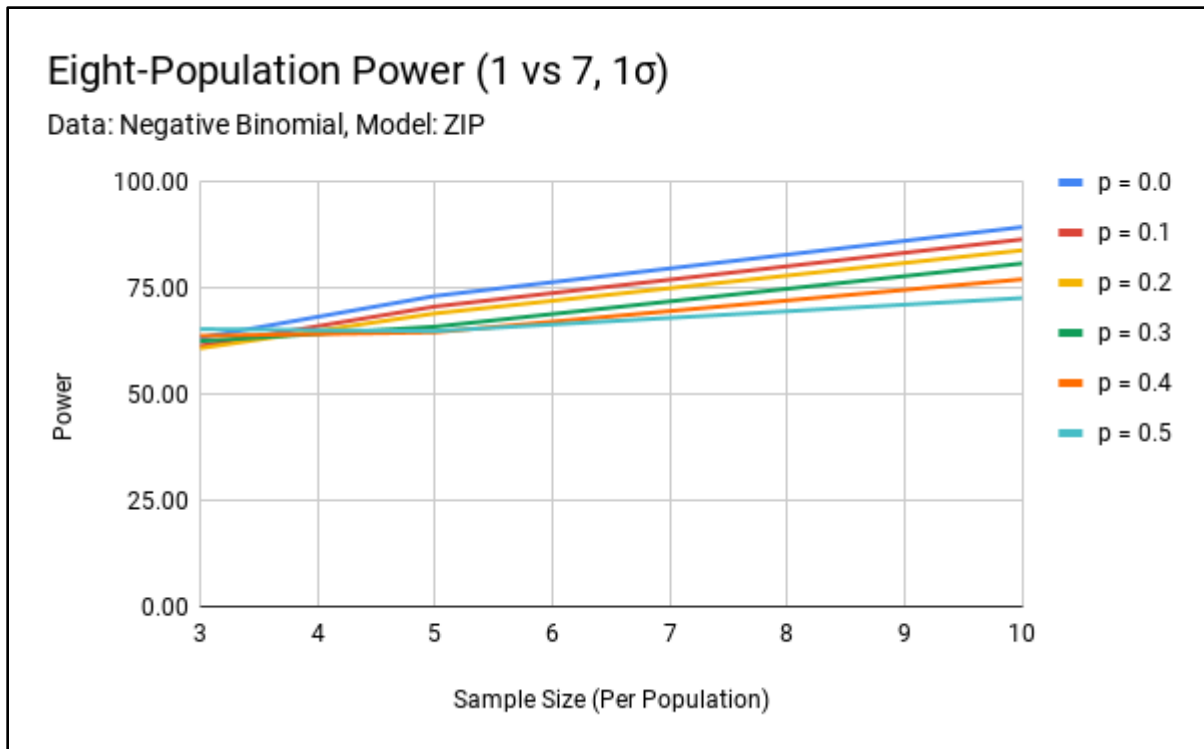


Figure B27. Negative Binomial Sample – ZIP Model Eight-Population Power (1 vs 7, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 24$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 44.36 | 8.46 | 6.80 | 6.36 | 6.59 | 5.66 |
| n = 5 | 79.16 | 11.22 | 7.62 | 6.65 | 6.70 | 5.97 |
| n = 10 | 99.66 | 22.08 | 11.13 | 8.31 | 7.11 | 6.40 |

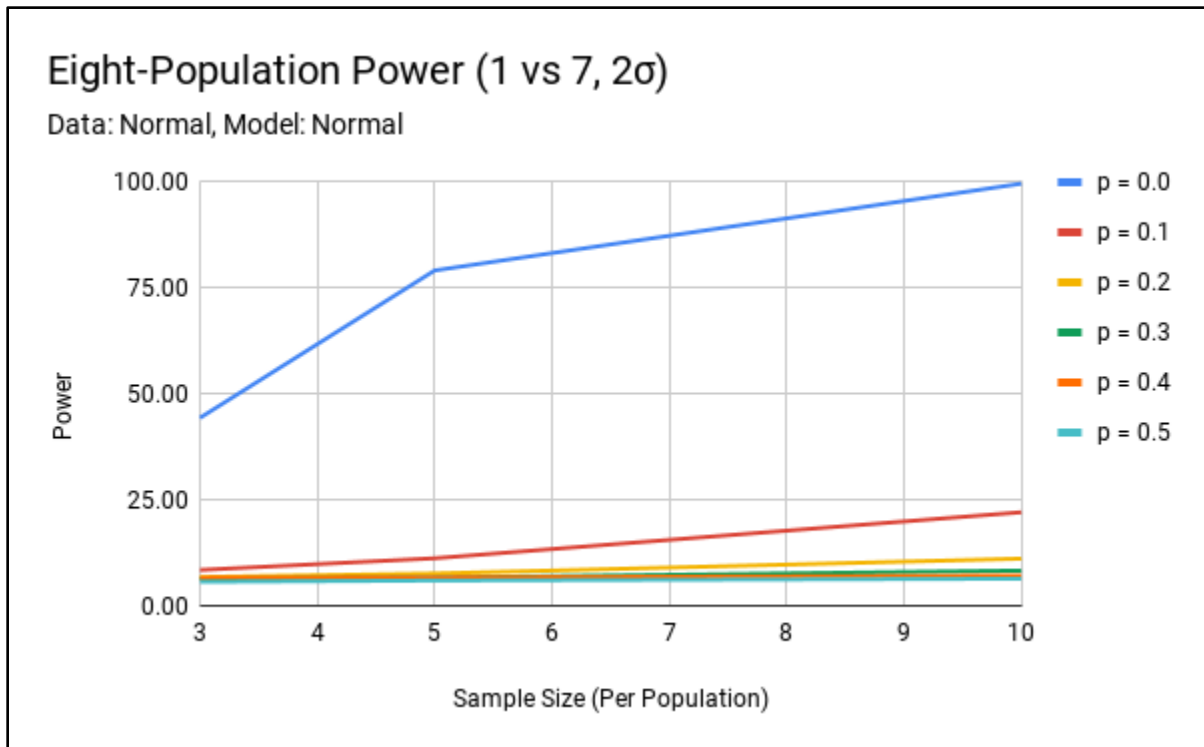


Figure B28. Normal Sample – Normal Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 24$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 61.05 | 7.48 | 6.21 | 12.29 | 14.45 | 12.26 |
| n = 5 | 85.26 | 7.91 | 4.41 | 4.84 | 8.04 | 11.38 |
| n = 10 | 99.49 | 17.76 | 8.74 | 6.43 | 5.43 | 5.68 |

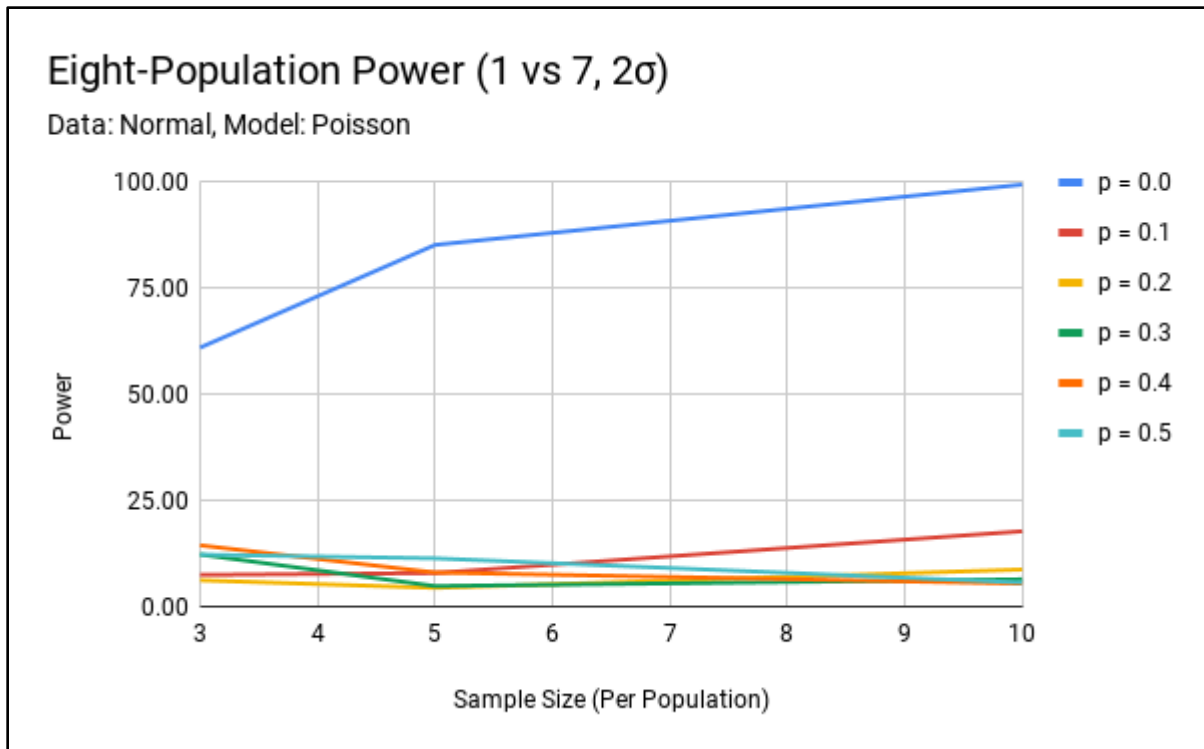


Figure B29. Normal Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = 24$

$\mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 0.00 | 0.36 | 0.91 | 1.09 | 2.05 | 1.91 |
| n = 5 | 0.00 | 0.00 | 0.25 | 0.97 | 1.21 | 1.46 |
| n = 10 | 0.82 | 0.51 | 0.25 | 0.12 | 0.13 | 0.68 |

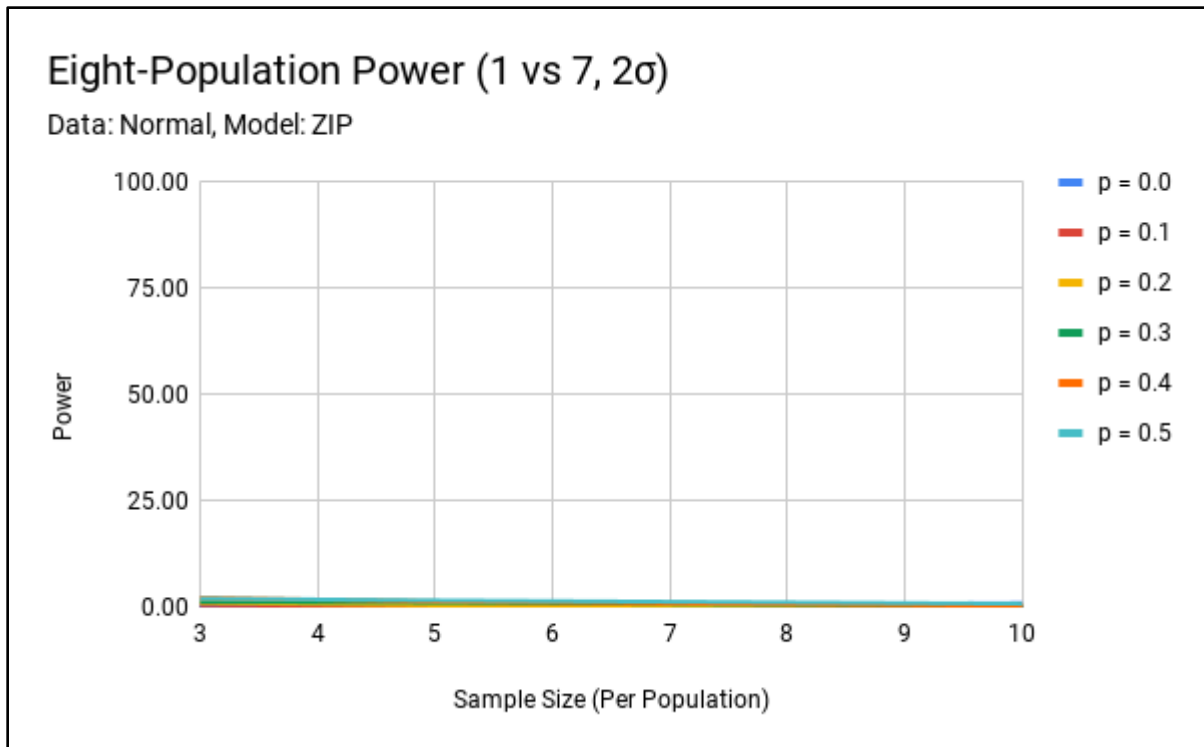


Figure B30. Normal Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 29$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 41.67 | 18.87 | 11.50 | 10.24 | 8.53 | 7.33 |
| n = 5 | 74.76 | 32.74 | 18.65 | 12.90 | 10.11 | 8.15 |
| n = 10 | 98.54 | 53.59 | 31.10 | 19.85 | 14.78 | 11.30 |

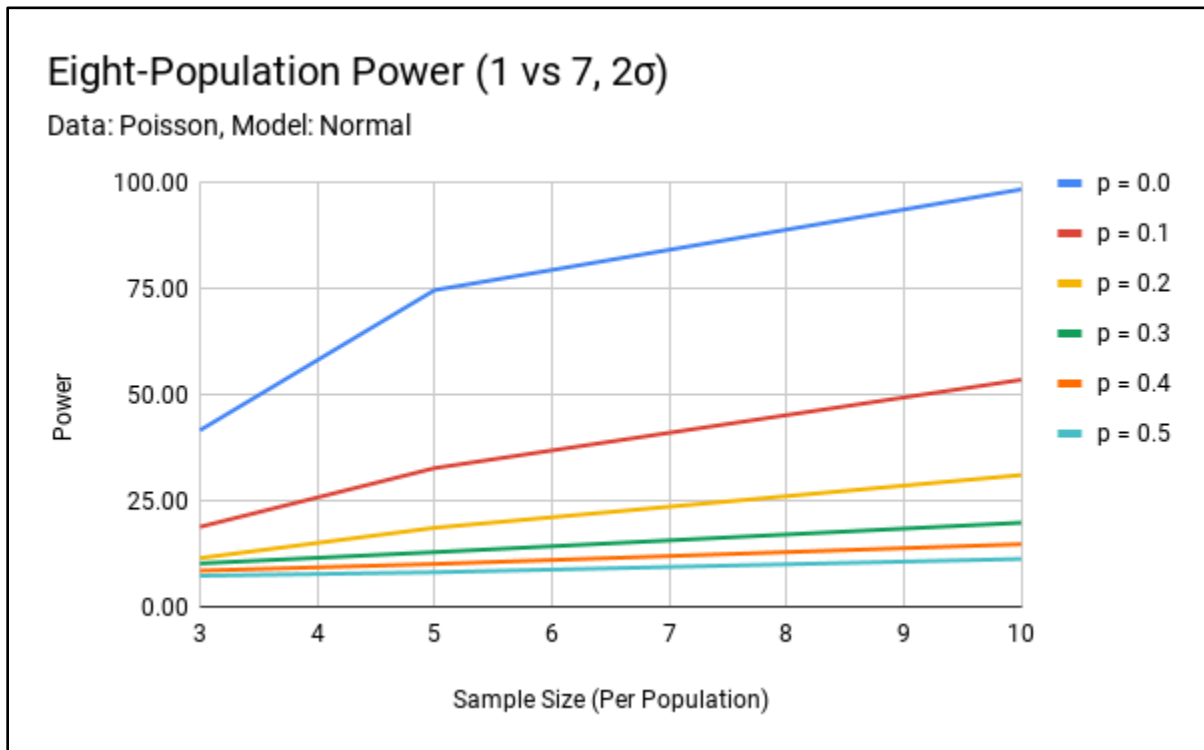


Figure B31. Poisson Sample – Normal Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 29$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 57.52 | 18.98 | 11.90 | 13.86 | 16.57 | 16.14 |
| n = 5 | 80.18 | 30.23 | 15.23 | 10.57 | 11.39 | 12.67 |
| n = 10 | 98.57 | 52.24 | 30.06 | 18.61 | 13.53 | 10.63 |

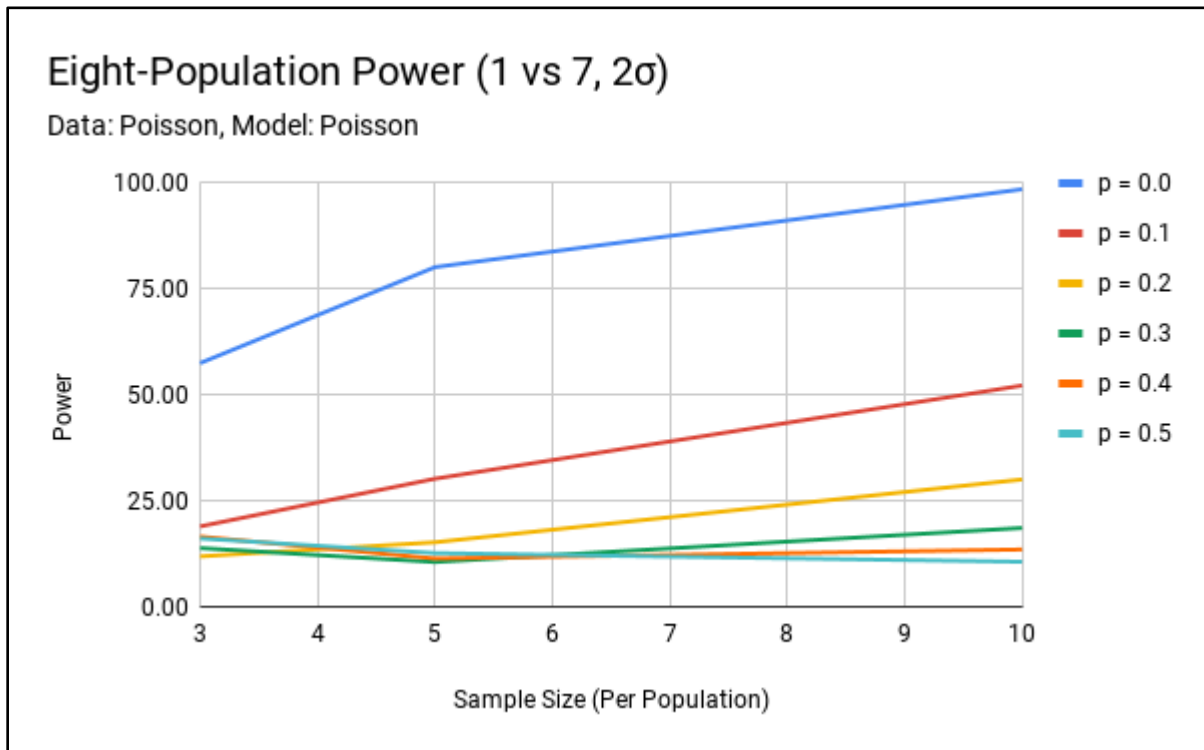


Figure B32. Poisson Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = 29$

$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 56.14 | 52.30 | 47.58 | 45.74 | 46.36 | 44.99 |
| n = 5 | 82.03 | 76.46 | 70.44 | 63.18 | 55.99 | 52.05 |
| n = 10 | 98.92 | 97.63 | 95.74 | 92.21 | 86.05 | 78.56 |

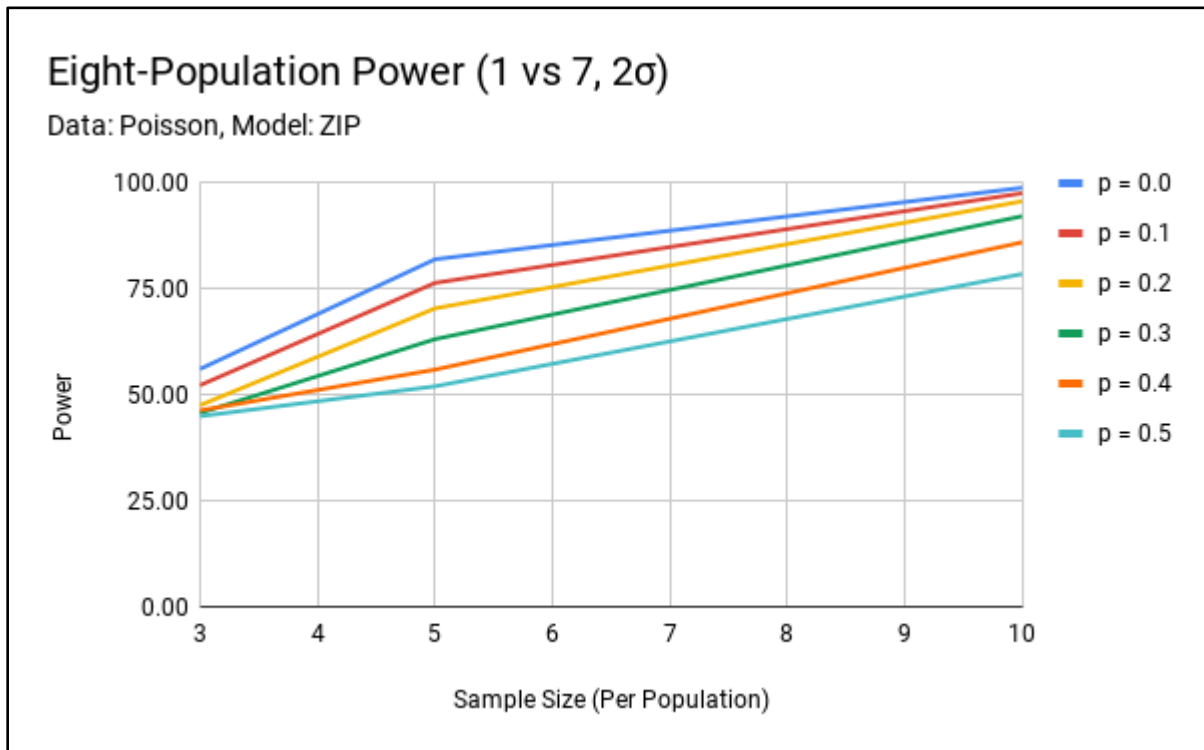


Figure B33. Poisson Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 32.6$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 40.66 | 23.02 | 16.90 | 16.80 | 13.62 | 9.01 |
| n = 5 | 71.59 | 38.90 | 24.70 | 19.82 | 15.25 | 10.26 |
| n = 10 | 97.91 | 66.35 | 44.10 | 30.25 | 21.84 | 15.82 |

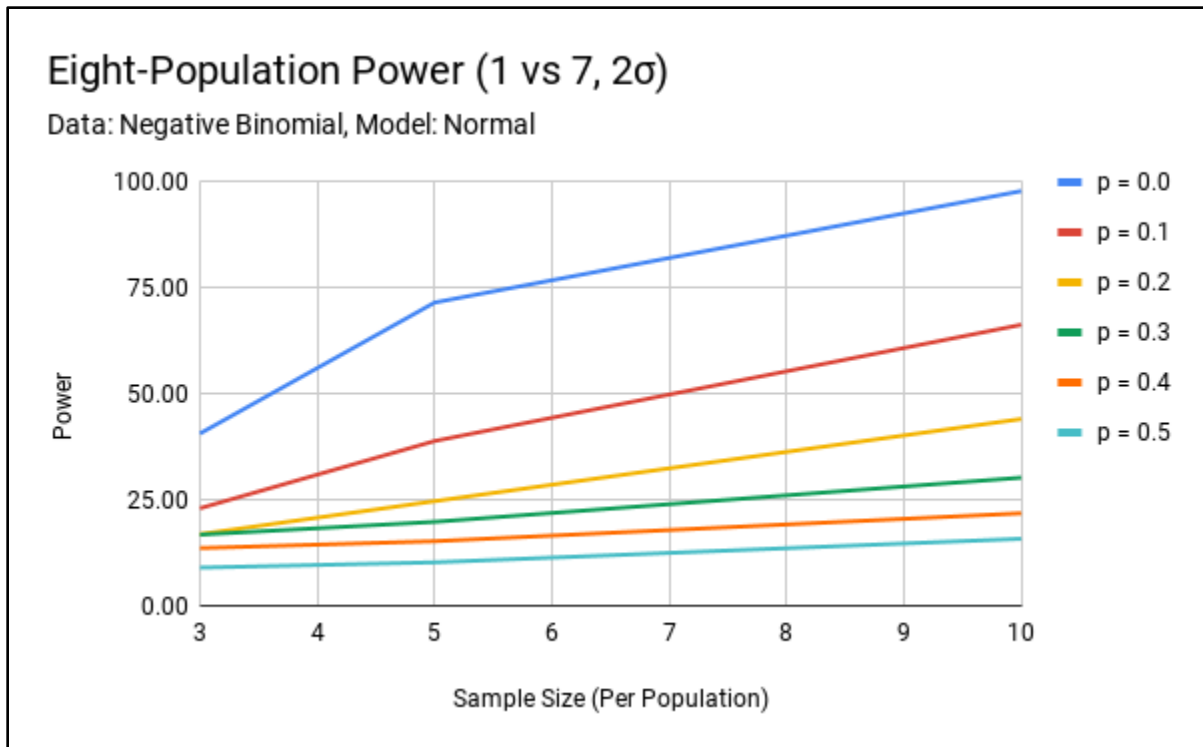


Figure B34. Negative Binomial Sample – Normal Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 32.6$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 54.89 | 25.89 | 17.60 | 18.11 | 19.45 | 18.90 |
| n = 5 | 76.93 | 40.39 | 24.16 | 17.47 | 15.39 | 15.56 |
| n = 10 | 98.03 | 65.76 | 42.77 | 29.17 | 20.85 | 16.13 |

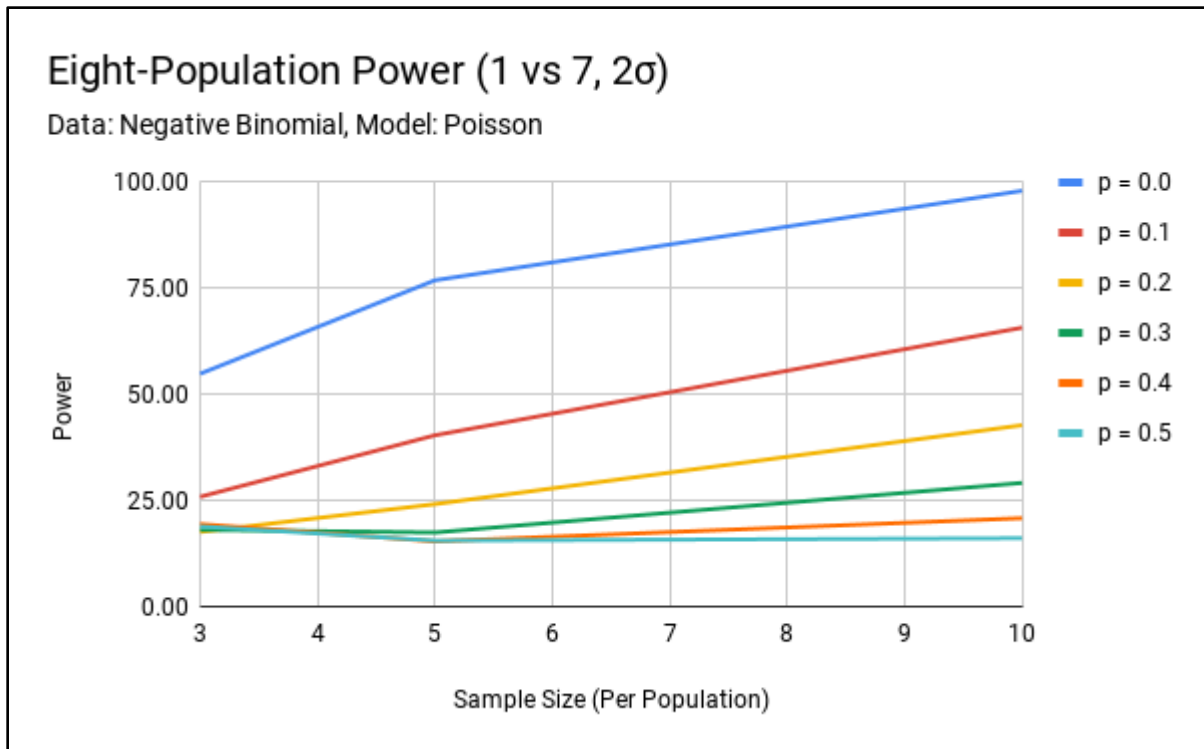


Figure B35. Negative Binomial Sample – Poisson Model Eight-Population Power (1 vs 7, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = 32.6$

$k(nb)_2 = k(nb)_3 = k(nb)_4 = k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 90.47 | 87.69 | 85.43 | 83.16 | 82.91 | 81.48 |
| n = 5 | 97.75 | 96.42 | 94.46 | 91.99 | 89.70 | 86.37 |
| n = 10 | 99.96 | 99.87 | 99.66 | 99.21 | 98.02 | 96.23 |

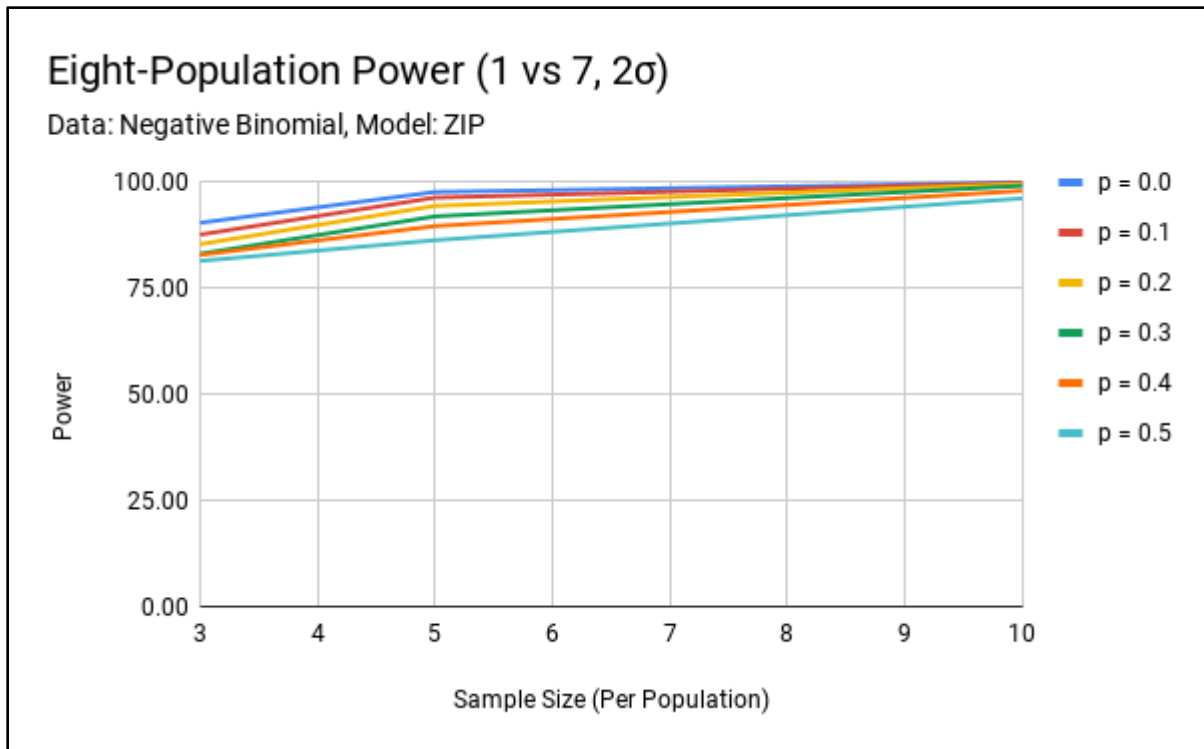


Figure B36. Negative Binomial Sample – ZIP Model Eight-Population Power (1 vs 7, 2σ).

Eight-Population 4 vs 4 Scenario

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 22$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 25.99 | 6.18 | 5.97 | 5.98 | 5.68 | 6.34 |
| n = 5 | 50.23 | 8.03 | 5.82 | 5.85 | 5.81 | 5.98 |
| n = 10 | 89.35 | 11.35 | 7.54 | 6.73 | 6.03 | 5.56 |

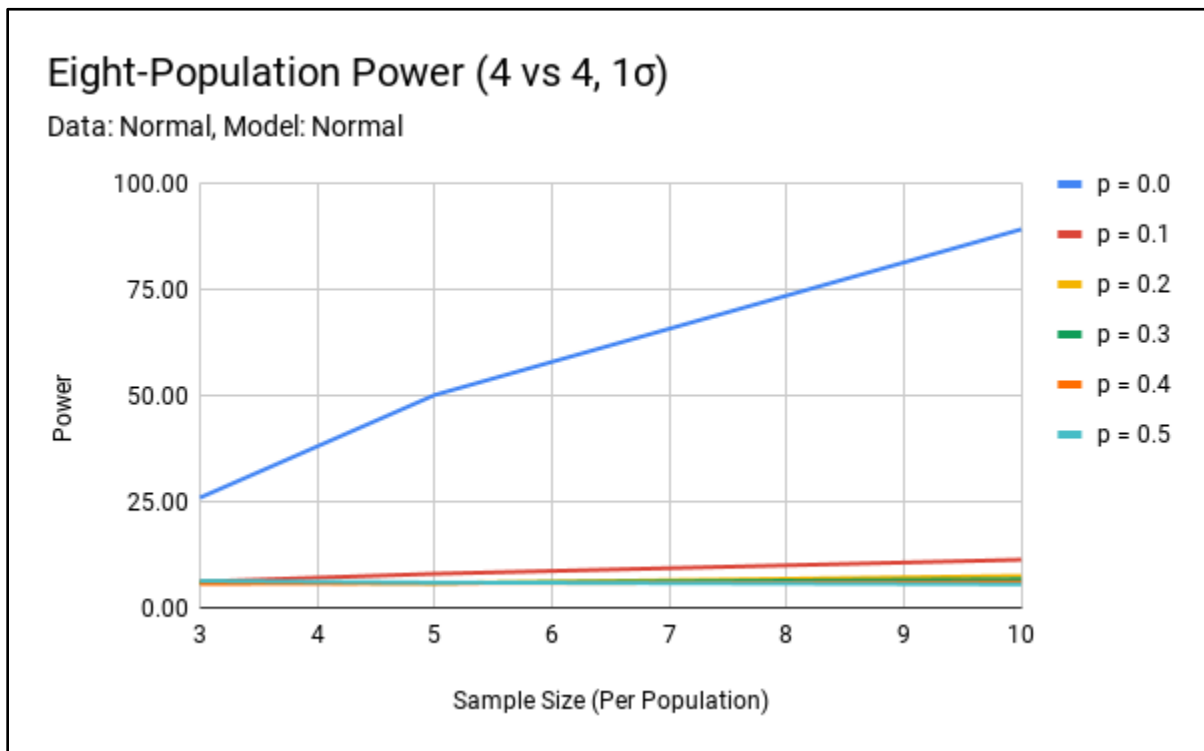


Figure B37. Normal Sample – Normal Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 22$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 42.21 | 5.64 | 5.79 | 12.15 | 15.04 | 13.34 |
| n = 5 | 60.96 | 5.49 | 3.80 | 4.24 | 7.67 | 11.05 |
| n = 10 | 91.62 | 9.64 | 5.83 | 5.22 | 5.37 | 5.33 |

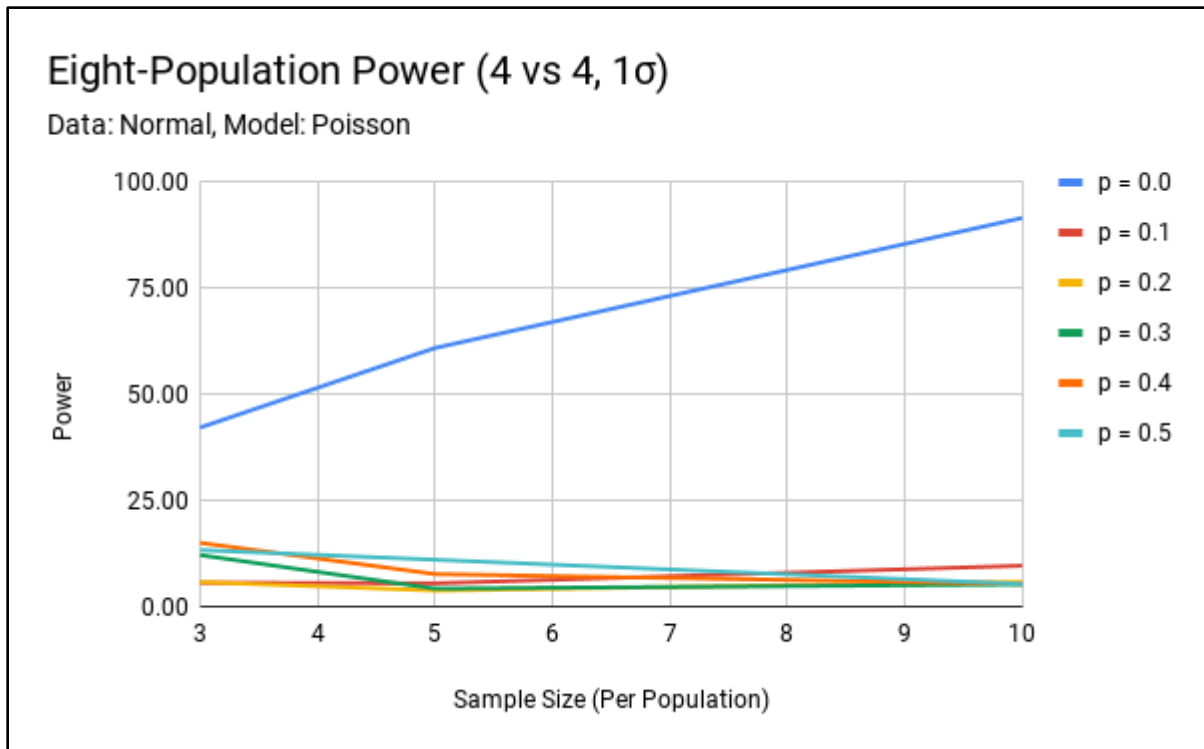


Figure B38. Normal Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 22$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 0.00 | 0.24 | 0.50 | 0.87 | 1.39 | 1.33 |
| n = 5 | 0.00 | 0.00 | 0.17 | 0.69 | 0.79 | 1.19 |
| n = 10 | 0.01 | 0.01 | 0.00 | 0.01 | 0.05 | 0.51 |

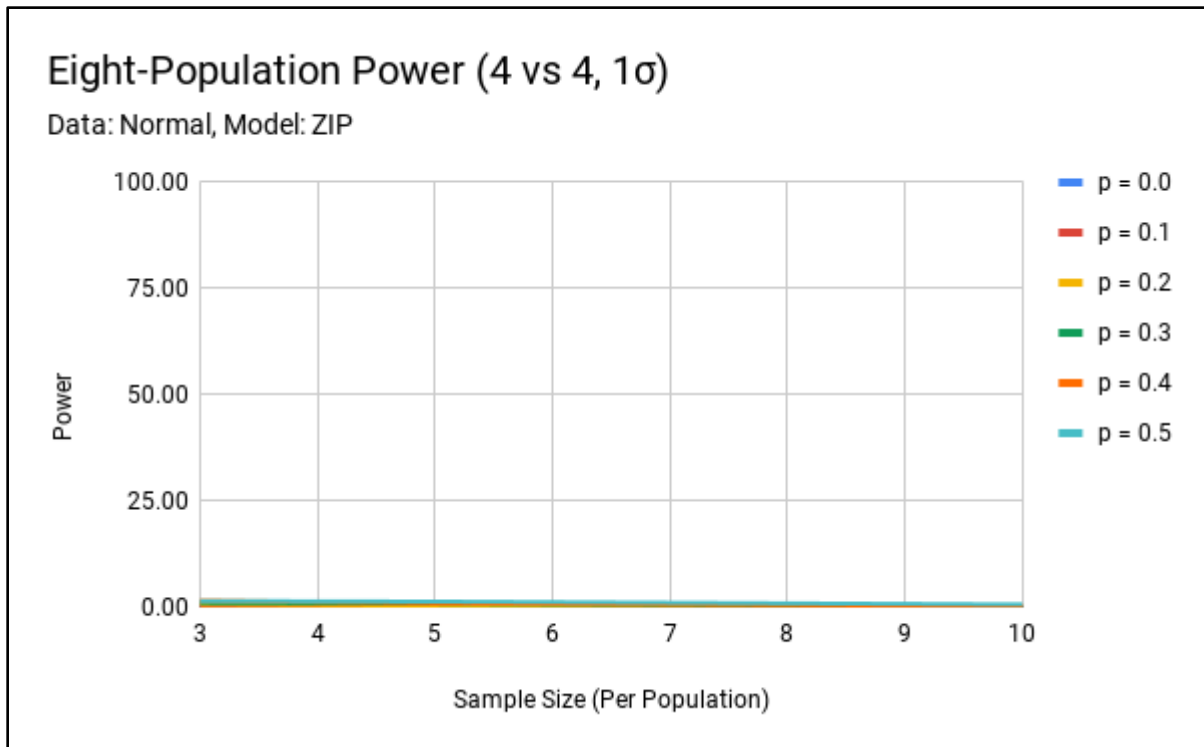


Figure B39. Normal Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 24.5$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 23.95 | 10.07 | 7.50 | 7.03 | 6.52 | 6.63 |
| n = 5 | 45.73 | 16.14 | 10.38 | 8.14 | 7.34 | 6.91 |
| n = 10 | 86.00 | 31.08 | 16.39 | 12.02 | 9.28 | 7.45 |

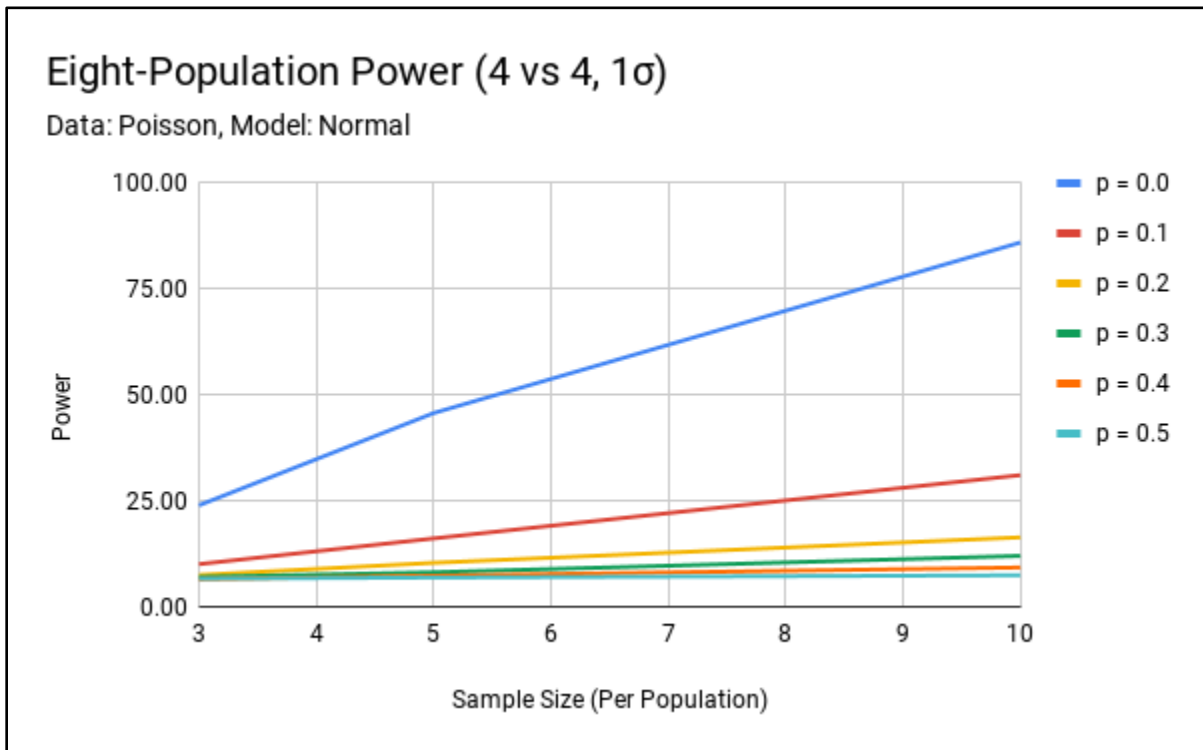


Figure B40. Poisson Sample – Normal Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 24.5$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 22.20 | 12.36 | 9.64 | 8.06 | 7.01 | 6.69 |
| n = 5 | 44.37 | 20.43 | 13.20 | 9.90 | 8.23 | 6.89 |
| n = 10 | 83.98 | 41.95 | 24.47 | 16.21 | 11.93 | 9.69 |

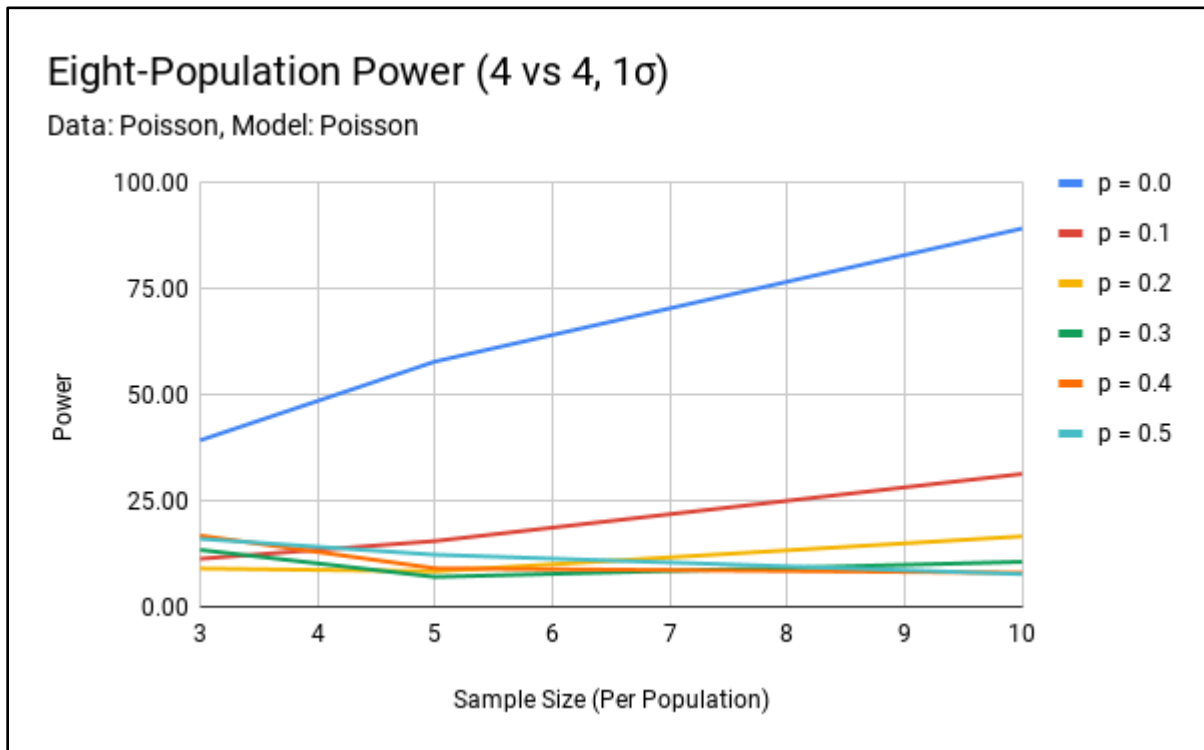


Figure B41. Poisson Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 24.5$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 33.92 | 31.54 | 31.14 | 32.78 | 33.15 | 34.30 |
| n = 5 | 57.69 | 50.43 | 46.37 | 40.49 | 37.18 | 38.21 |
| n = 10 | 89.83 | 85.57 | 80.93 | 73.25 | 66.50 | 57.65 |

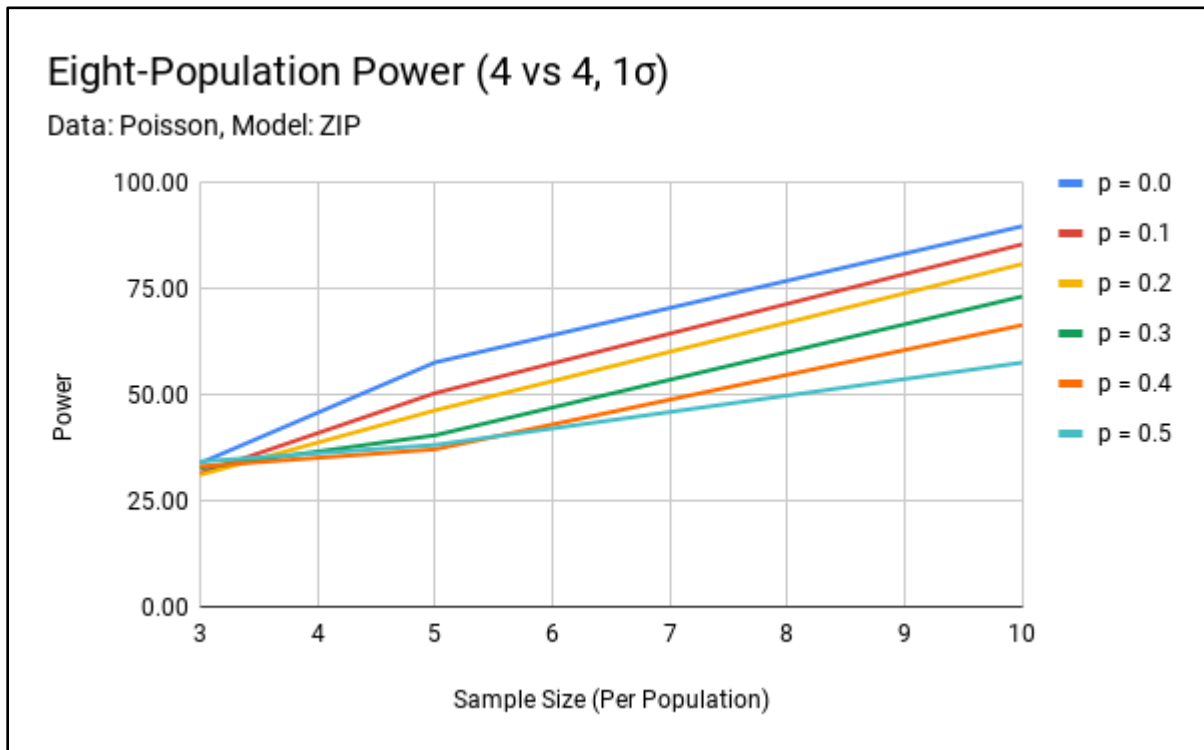


Figure B42. Poisson Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 26.3$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 22.20 | 12.36 | 9.64 | 8.06 | 7.01 | 6.69 |
| n = 5 | 44.37 | 20.43 | 13.20 | 9.90 | 8.23 | 6.89 |
| n = 10 | 83.98 | 41.95 | 24.47 | 16.21 | 11.93 | 9.69 |

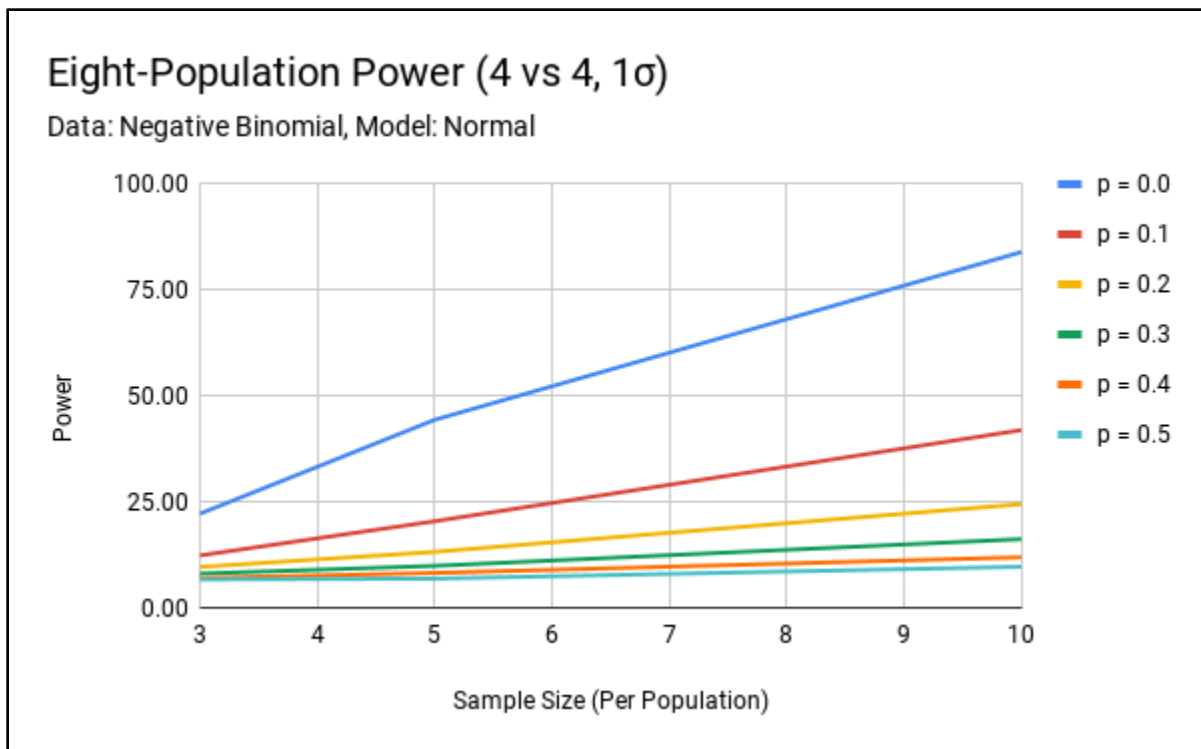


Figure B43. Negative Binomial Sample – Normal Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 26.3$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 38.80 | 15.96 | 12.42 | 14.81 | 17.58 | 18.48 |
| n = 5 | 55.91 | 23.10 | 13.32 | 10.80 | 11.22 | 13.45 |
| n = 10 | 87.27 | 44.61 | 24.61 | 16.38 | 12.14 | 10.32 |

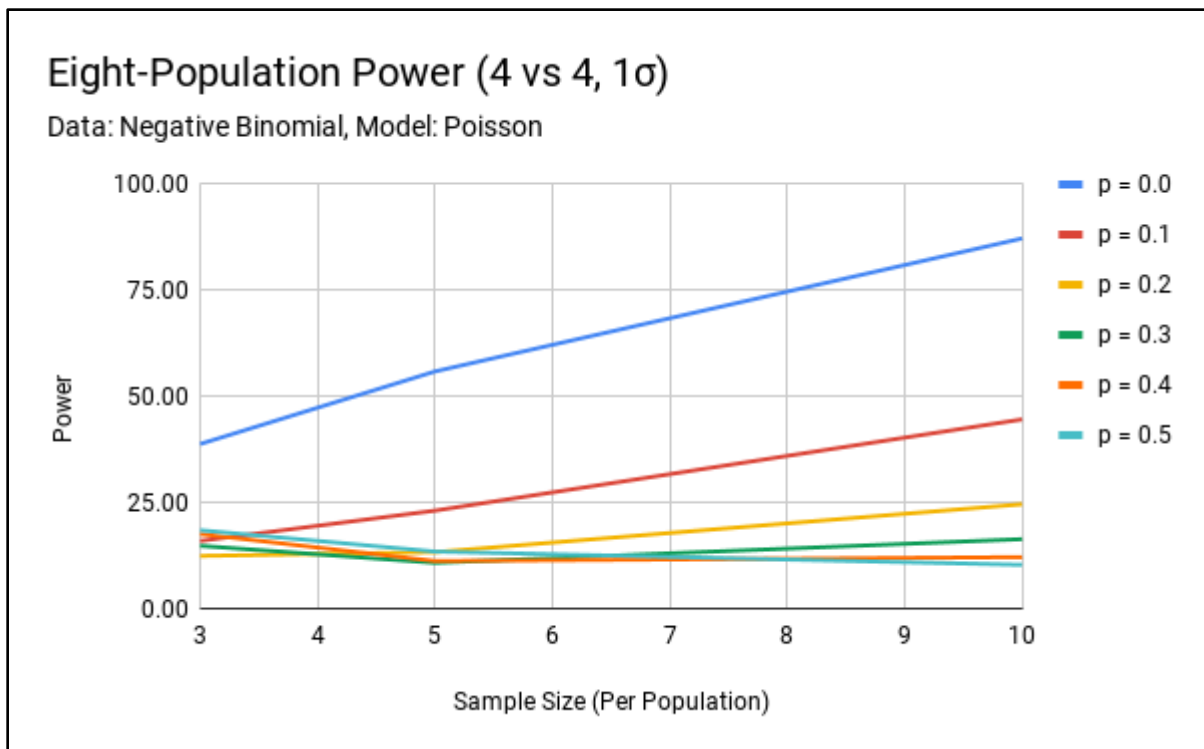


Figure B44. Negative Binomial Sample – Poisson Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 1σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 26.3$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 80.60 | 78.53 | 76.03 | 75.81 | 75.29 | 74.39 |
| n = 5 | 91.29 | 89.64 | 87.05 | 84.06 | 81.06 | 79.42 |
| n = 10 | 99.18 | 98.66 | 97.62 | 96.45 | 93.74 | 90.98 |

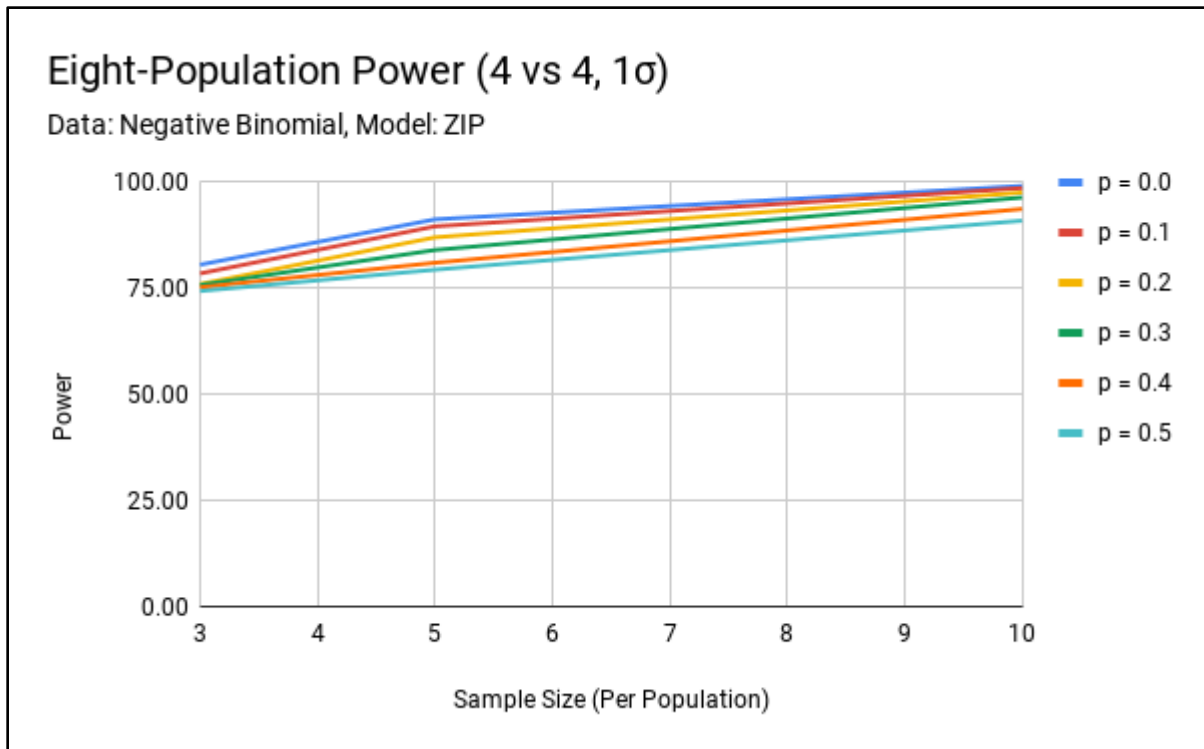


Figure B45. Negative Binomial Sample – ZIP Model Eight-Population Power (4 vs 4, 1σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 24$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 84.83 | 15.68 | 7.41 | 6.77 | 6.35 | 6.84 |
| n = 5 | 99.53 | 19.13 | 10.34 | 7.52 | 7.03 | 6.07 |
| n = 10 | 100.00 | 35.41 | 16.59 | 10.70 | 8.70 | 7.70 |

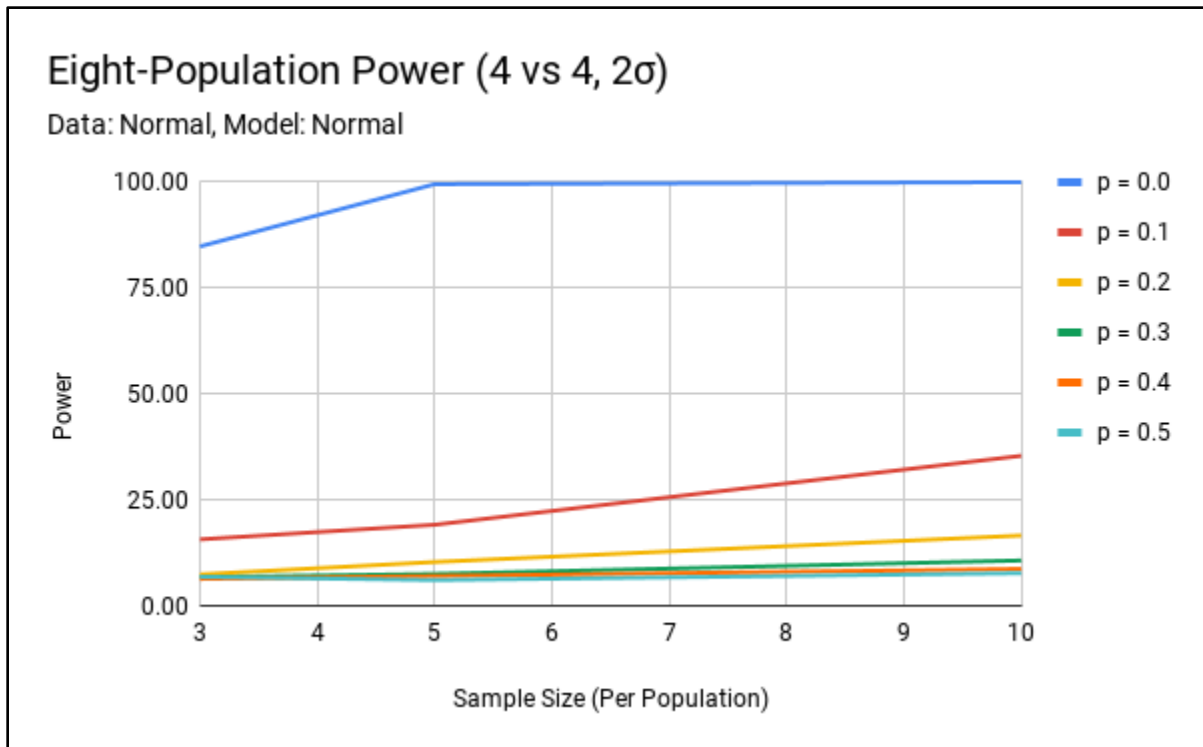


Figure B46. Normal Sample – Normal Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 24$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 93.75 | 12.53 | 8.10 | 13.06 | 16.25 | 15.79 |
| n = 5 | 99.75 | 15.39 | 7.10 | 5.88 | 8.29 | 11.00 |
| n = 10 | 100.00 | 68.85 | 33.77 | 16.53 | 8.27 | 6.51 |

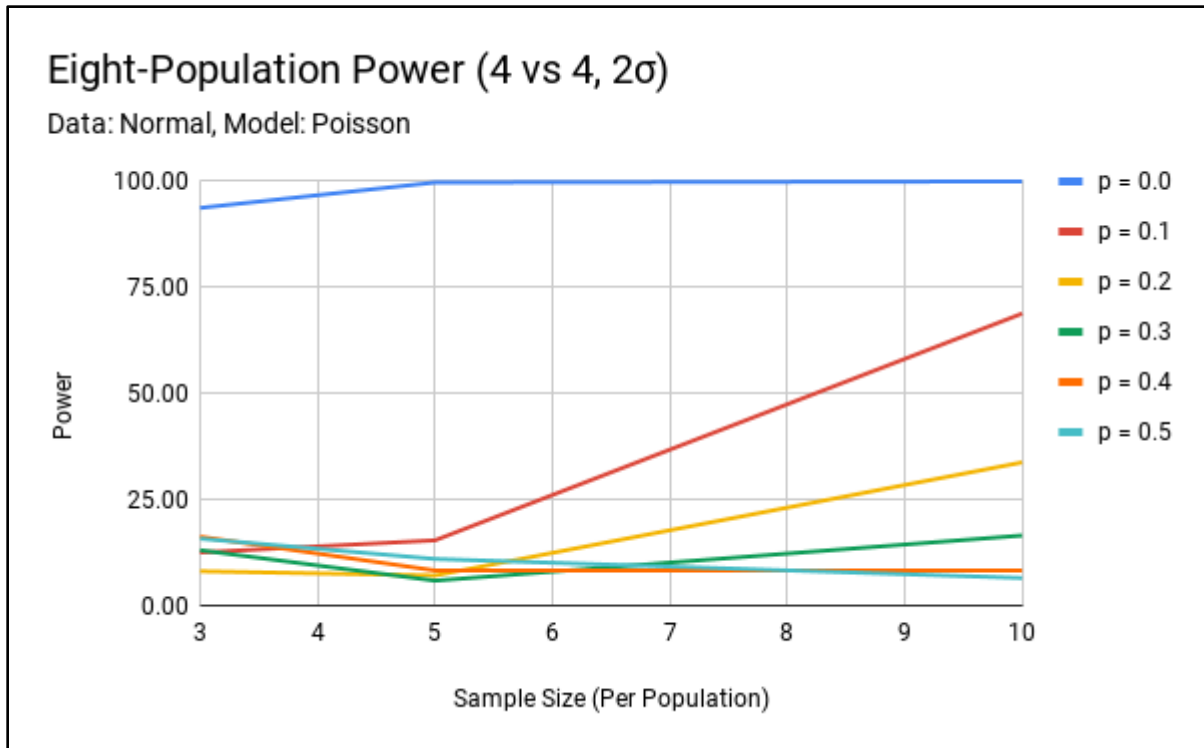


Figure B47. Normal Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Normal

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 24$

$\mu_5 = \mu_6 = \mu_7 = \mu_8 = 20$

$\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = \sigma_8 = 2$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 0.04 | 0.31 | 2.08 | 2.70 | 3.10 | 3.25 |
| n = 5 | 1.84 | 0.91 | 0.56 | 1.47 | 3.04 | 3.34 |
| n = 10 | 44.94 | 29.20 | 16.89 | 8.06 | 3.30 | 1.49 |

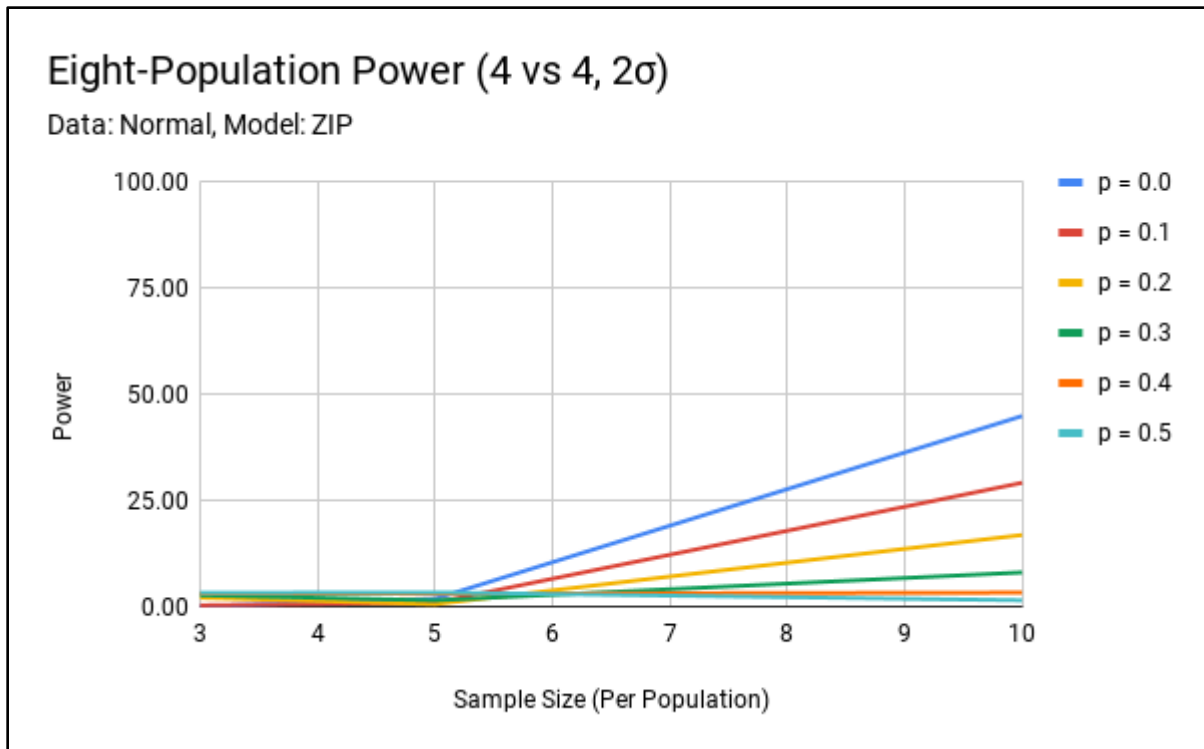


Figure B48. Normal Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 29$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 75.49 | 28.47 | 15.39 | 11.80 | 9.66 | 7.74 |
| n = 5 | 98.07 | 46.35 | 24.39 | 16.21 | 12.69 | 9.58 |
| n = 10 | 100.00 | 80.45 | 50.72 | 30.95 | 20.82 | 14.99 |

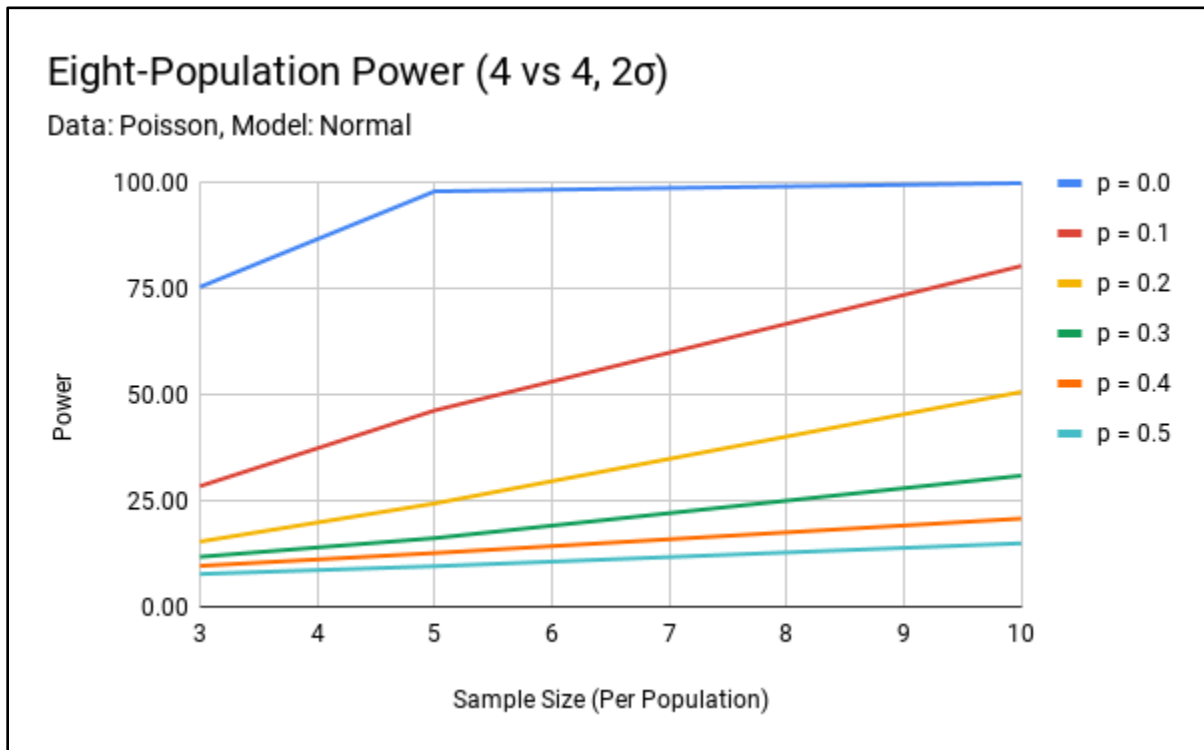


Figure B49. Poisson Sample – Normal Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 29$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 88.42 | 30.29 | 15.69 | 16.41 | 19.26 | 21.20 |
| n = 5 | 99.21 | 49.75 | 24.08 | 15.43 | 13.67 | 15.67 |
| n = 10 | 100.00 | 83.82 | 52.86 | 32.84 | 21.14 | 15.93 |

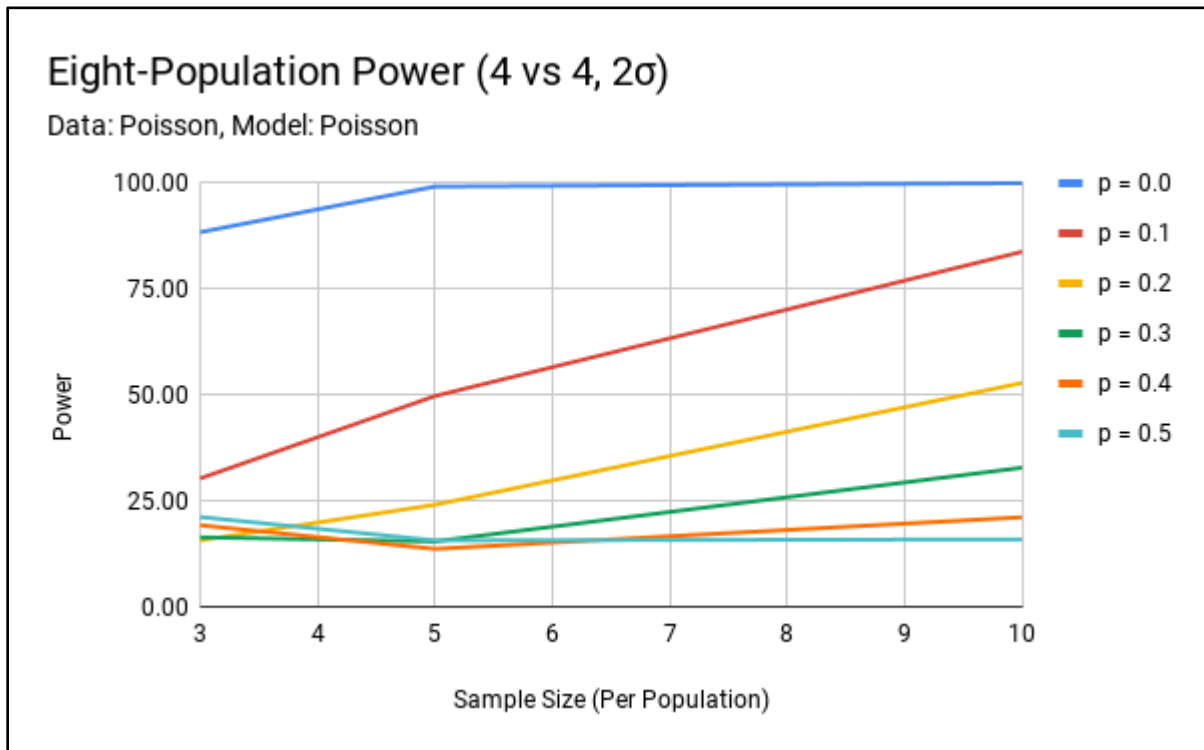


Figure B50. Poisson Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Poisson

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 29$

$\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 20$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 92.49 | 89.30 | 85.22 | 81.93 | 79.12 | 74.92 |
| n = 5 | 99.61 | 99.15 | 98.10 | 95.60 | 92.49 | 87.82 |
| n = 10 | 100.00 | 100.00 | 99.99 | 99.98 | 99.89 | 99.36 |

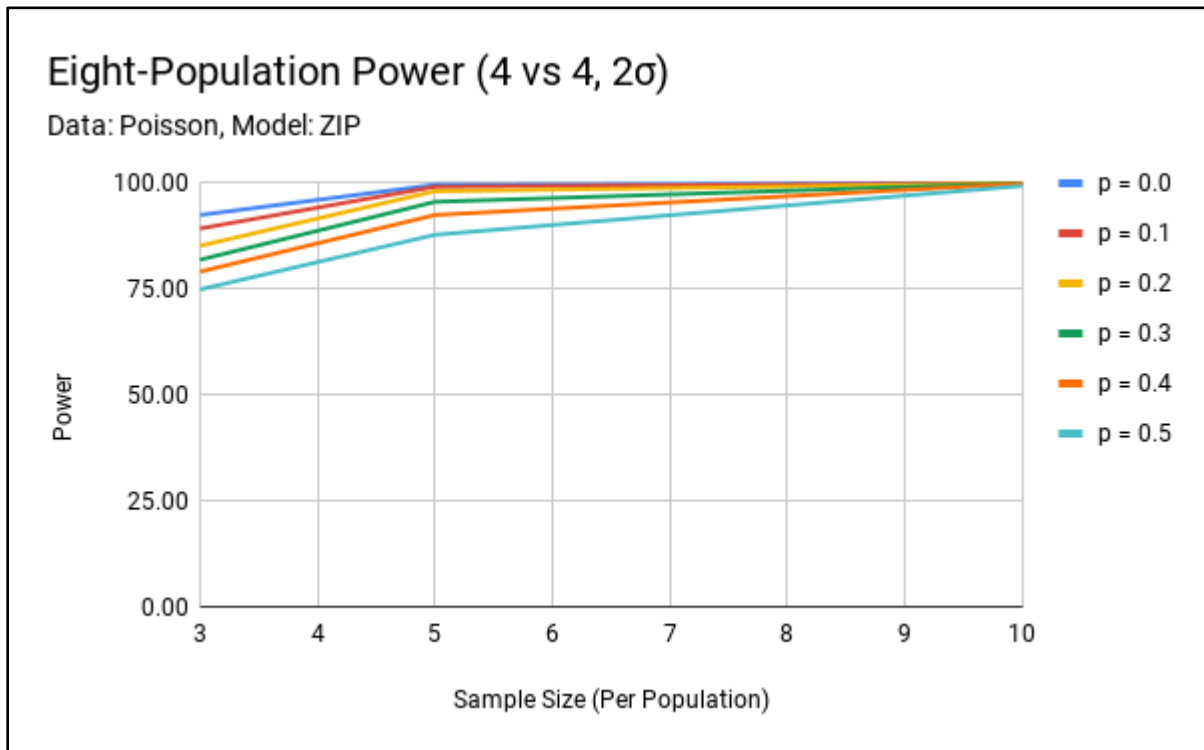


Figure B51. Poisson Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 32.6$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Normal | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|---------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 71.47 | 34.40 | 20.22 | 14.34 | 11.84 | 9.91 |
| n = 5 | 97.15 | 57.84 | 34.17 | 22.89 | 16.05 | 12.07 |
| n = 10 | 100.00 | 91.44 | 68.79 | 47.45 | 31.43 | 21.69 |

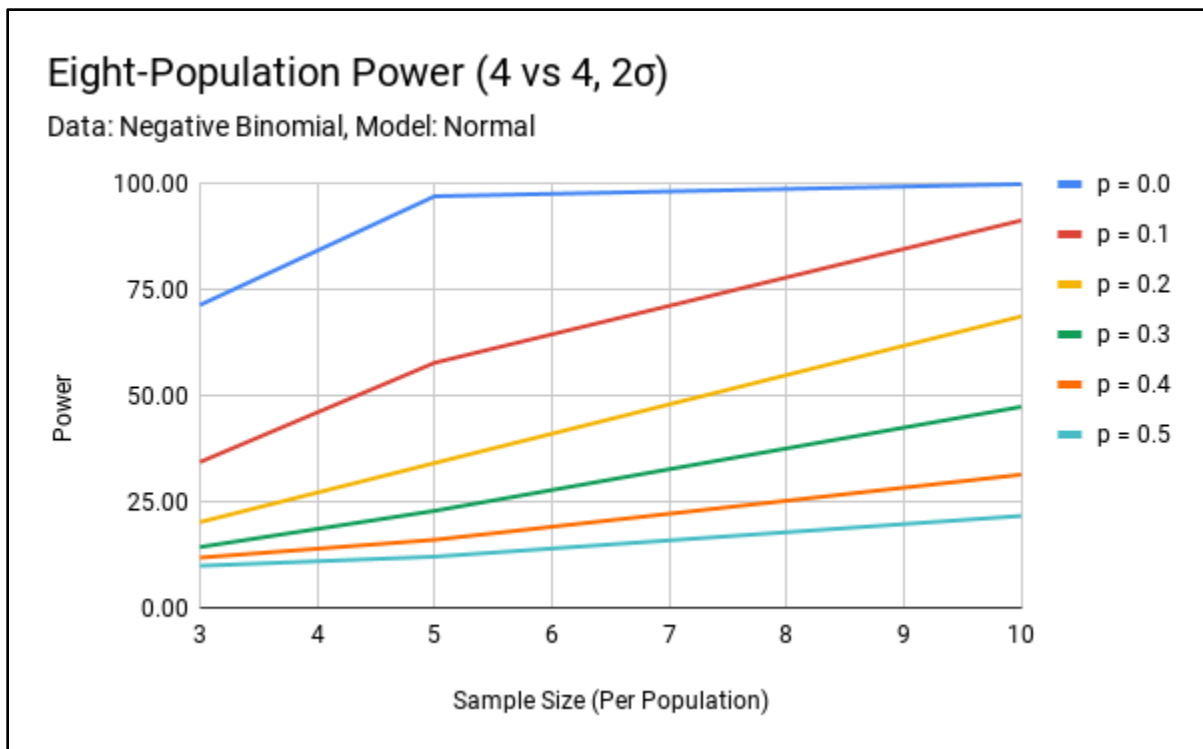


Figure B52. Negative Binomial Sample – Normal Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 32.6$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: Poisson | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|----------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 87.17 | 41.27 | 24.35 | 20.82 | 22.90 | 24.21 |
| n = 5 | 98.53 | 64.33 | 38.51 | 24.39 | 19.51 | 19.61 |
| n = 10 | 100.00 | 93.50 | 72.18 | 50.85 | 34.83 | 24.41 |

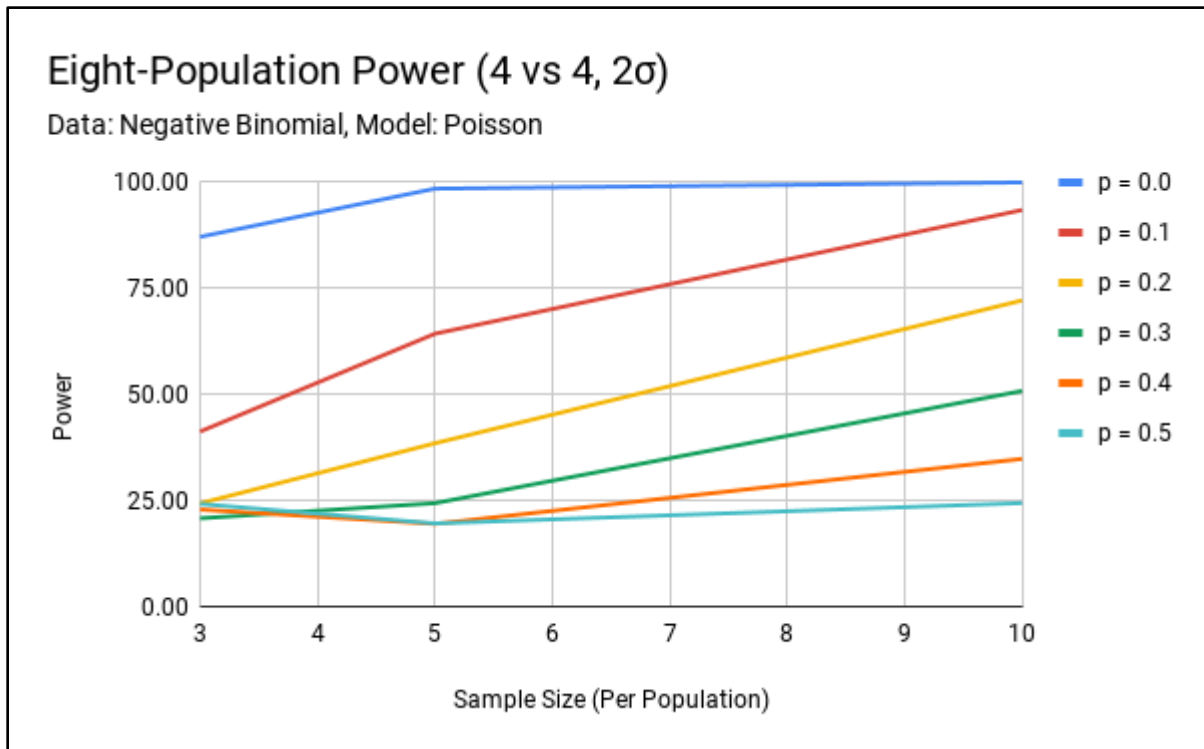


Figure B53. Negative Binomial Sample – Poisson Model Eight-Population Power (4 vs 4, 2σ).

Sample distribution: Negative Binomial

Simulations: 10,000

Effect size: 2σ

$n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = n_7 = n_8$

$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8$

$k(nb)_1 = k(nb)_2 = k(nb)_3 = k(nb)_4 = 32.6$

$k(nb)_5 = k(nb)_6 = k(nb)_7 = k(nb)_8 = 20$

$p(nb)_1 = p(nb)_2 = p(nb)_3 = p(nb)_4 = p(nb)_5 = p(nb)_6 = p(nb)_7 = p(nb)_8 = 0.5$

| Model: ZIP | p = 0.0 | p = 0.1 | p = 0.2 | p = 0.3 | p = 0.4 | p = 0.5 |
|------------|---------|---------|---------|---------|---------|---------|
| n = 3 | 99.63 | 99.04 | 98.28 | 97.58 | 96.60 | 95.58 |
| n = 5 | 99.98 | 99.97 | 99.92 | 99.66 | 99.43 | 98.60 |
| n = 10 | 100.00 | 100.00 | 100.00 | 100.00 | 99.98 | 99.97 |

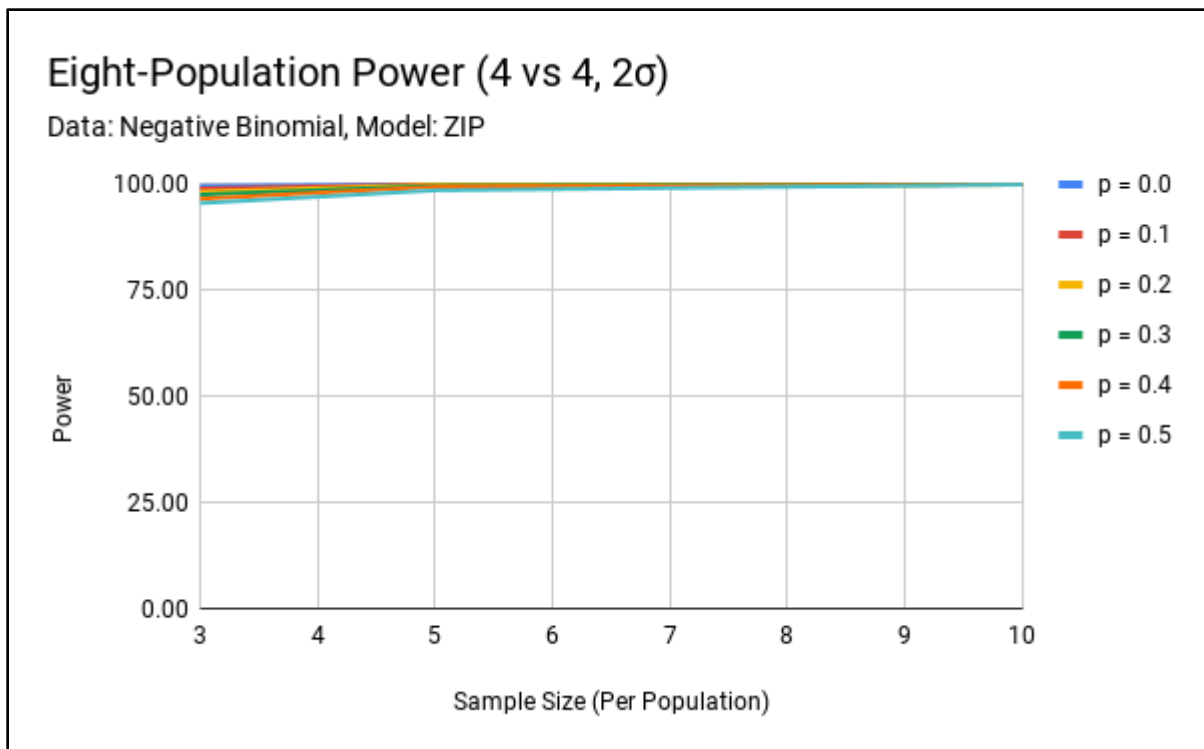


Figure B54. Negative Binomial Sample – ZIP Model Eight-Population Power (4 vs 4, 2σ).

APPENDIX C. EIGHT-POPULATION SAS CODE

Normal Samples

```
*****;
/*

Simulation Parameters-----

    pop = population ID
    p = percentage of zero values in population
    n = sample size for population
    mu = mean of normal distribution for population
    sigma = standard deviation of normal dist for population
    samples = number of samples to simulate

*/
*****;

options threads cpubcount=2;
options mprint symbolgen formchar="|----|+|----+=|-\<>";

libname lyzn 'C:\Users\Lucas Young\Desktop\Thesis\Data';

%let t0 = %sysfunc(datetime()); *** Collect timing data. ***;

ods html close;
*ods html;
dm 'log; clear; output; clear;';
ods html path='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods html gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods listing gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';

*ods trace on / listing;

%macro mix_z_Norm(pop,p,n,mu,sigma,samples);
%global gp&pop gn&pop gmu&pop gsigma&pop gsamples&pop;
%let gp&pop          = %sysfunc(compress(&p,));
%let gn&pop          = &n;
%let gmu&pop         = &mu;
%let gsigma&pop      = &sigma;
%let gsamples&pop   = &samples;

data gen_data_&pop;
```

```

pop=&pop;          *** Population ID          ***;
p=&p;              *** Proportion of zeroes in pop ***;
n=&n;              *** Sample size from pop 1. ***;
mu=&mu;           *** Mean for Normal from pop ***;
sigma=&sigma;      *** SD for normal from pop ***;
samples=&samples; *** Number of samples from pop ***;

call streaminit(0);

do sample=1 to &samples;
do i=1 to &n;
    D=rand('bernoulli',&p); *** Use Bernoulli to select dist. ***;
    if D=0 then Y=rand('normal',&mu,&sigma);
    else if D=1 then Y=0;
output;
end;
end;
run;

%mend mix_z_Norm;

%macro ODSOff();
ods graphics off; ods exclude all; ods noresults;
options nonotes nosource nosource2 errors=0;
%mend;

%macro ODSOn();
ods graphics on; ods exclude none; ods results;
options notes source source2 errors=20;
%mend;

%mix_z_Norm(1, 0.1, 3, 24, 2, 10000);
%mix_z_Norm(2, 0.1, 3, 24, 2, 10000);
%mix_z_Norm(3, 0.1, 3, 24, 2, 10000);
%mix_z_Norm(4, 0.1, 3, 24, 2, 10000);

%mix_z_Norm(5, 0.1, 3, 20, 2, 10000);
%mix_z_Norm(6, 0.1, 3, 20, 2, 10000);
%mix_z_Norm(7, 0.1, 3, 20, 2, 10000);
%mix_z_Norm(8, 0.1, 3, 20, 2, 10000);

/* Combine Samples */
data all;

```

```

set
  gen_data_1
  gen_data_2
  gen_data_3
  gen_data_4
  gen_data_5
  gen_data_6
  gen_data_7
  gen_data_8;
run;

/* Check small number of samples */
proc print data=all(obs=20);
run;

proc sort data=all;
  by sample pop;
run;

*** Run UNIVARIATE to check dist initially. ***;
*** If OK, then comment off. ***;

ods graphics on;
ods exclude ExtremeObs TestsForLocation BasicMeasures;
proc univariate plot normal data=all;
  where Y between 0 and 50;
  *class pop;
  var Y;
  histogram Y / endpoints = 0 to 50 by 0.5;
  title2 'Check Distribution of Simulated Data';
  title3 'All Samples Combined';
  output out=Univariate_All mean=mean;
run;

%ODSOFF

*****;
***** Traditional ANOVA Approach *****;
*****;

ods output ModelANOVA = YANOVAscale;
proc genmod data=all;
  by sample;
  class pop;
  model Y = pop / dist=normal type3 noscale scale=pearson;
  lsmeans pop / diff;

```

```

title2 'Traditional ANOVA Approach';
run;

*****
*****      Poisson fit with Pearson Scaling      *****;
*****

ods output ModelANOVA = PoissonANOVA;
proc genmod data=all;
  by sample;
  class pop;
  model Y = pop / dist=poisson scale=pearson type3;
  lsmeans pop / diff;
  title2 'Generalized Linear Model Approach - Scaled Poisson Fit';
run;

*****
*****      Zero-Inflated Poisson Fit      *****;
*****

ods output ModelANOVA = ZIPANOVA;
proc genmod data=all;
  by Sample;
  class pop;
  model Y = pop / dist=zip type3;
  lsmeans pop / diff;
  zeromodel;
  title2 'Generalized Linear Model Approach - ZIP Fit';
run;

%ODSON;

proc format;
  value sig  0  -  .05 = 'Reject'
           .05 <- 1 = 'DNR';
run;

*****
*****      Rejection Rate for All Samples      *****;
*****

proc freq data=YANOVAscale;
  tables ProbF;
  format ProbF sig.;
  title2 'Rejection Rates for Treatment - ANOVA scale Y';
run;

```

```

proc freq data=PoissonANOVA;
    tables ProbChiSq;
    format ProbChiSq sig.;
    title2 'Rejection Rates for Treatment - Poisson w/ Scale=Pearson';
run;

proc freq data=ZIPANOVA;
    tables ProbChiSq;
    format ProbChiSq sig.;
    title2 'Rejection Rates for Treatment - ZIP Model';
run;

*****
*****                               Realized Effect Size                               *****
*****
proc transpose data=Univariate_all out=Univariate_all_T (drop=_LABEL_
rename=(_NAME_=Variable));
    title2 'Realized Effect Size';
run;

data Univariate_all_T;
    set Univariate_all_T;
    diff = COL1 - COL5;
run;

proc print data=Univariate_all_T;
    where Variable = 'mean';
run;

*****
*****                               Run Time                               *****
*****

%let t1 = %sysfunc(datetime());
%let elapsedTime = %sysevalf(&t1-&t0);
%put &elapsedTime;

```

Poisson Samples

```
*****;
/*
Simulation Parameters-----

    pop = population ID
    p = percentage of zero values in population
    n = sample size for population
    lambda = lambda of Poisson distribution
    samples = number of samples to simulate

*/
*****;

options threads cpucount=2;
options mprint symbolgen formchar="|----|+|----+=|-\<>";

libname lyzn 'C:\Users\Lucas Young\Desktop\Thesis\Data';

%let t0 = %sysfunc(datetime()); *** Collect timing data. ***;

ods html close;
*ods html;
dm 'log; clear; output; clear;';
ods html path='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods html gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods listing gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';

*ods trace on / listing;

%macro mix_z_poisson(pop,p,n,lambda,samples);
%global gp&pop gn&pop glambda&pop gsamples&pop;
%let gp&pop = %sysfunc(compress(&p,));
%let gn&pop = &n;
%let glambda&pop = &lambda;
%let gsamples&pop = &samples;

data gen_data_&pop;
    pop=&pop;          *** Population ID          ***;
    p=&p;              *** Proportion of zeroes in pop ***;
    n=&n;              *** Sample size from pop 1.      ***;
```

```

lambda=&lambda;    *** Lambda for Poisson distribution ***;
samples=&samples;  *** Number of samples from pop          ***;

call streaminit(0);

do sample=1 to &samples;
do i=1 to &n;
    D=rand('bernoulli',&p);    *** Use Bernoulli to select dist. ***;
    if D=0 then Y=rand('poisson',&lambda);
    else if D=1 then Y=0;
output;
end;
end;
run;

ods graphics on;

%mend mix_z_poisson;

%macro ODSOff();
ods graphics off; ods exclude all; ods noresults;
options nonotes nosource nosource2 errors=0;
%mend;

%macro ODSOn();
ods graphics on; ods exclude none; ods results;
options notes source source2 errors=20;
%mend;

%mix_z_poisson(1, 0.1, 3, 29, 10000);
%mix_z_poisson(2, 0.1, 3, 29, 10000);
%mix_z_poisson(3, 0.1, 3, 29, 10000);
%mix_z_poisson(4, 0.1, 3, 29, 10000);

%mix_z_poisson(5, 0.1, 3, 20, 10000);
%mix_z_poisson(6, 0.1, 3, 20, 10000);
%mix_z_poisson(7, 0.1, 3, 20, 10000);
%mix_z_poisson(8, 0.1, 3, 20, 10000);

/* Combine Samples */
data all;
set
gen_data_1
gen_data_2
gen_data_3
gen_data_4

```



```

*****
*****      Poisson fit with Pearson Scaling      *****;
*****

```

```

ods output ModelANOVA = PoissonANOVA;
proc genmod data=all;
  by sample;
  class pop;
  model Y = pop / dist=poisson scale=pearson type3;
  lsmeans pop / diff;
  output out=PearPReg pred=PearPPred;
  title2 'Generalized Linear Model Approach - Scaled Poisson Fit';
run;

```

```

*****
*****      Zero-Inflated Poisson Fit      *****;
*****

```

```

ods output ModelANOVA = ZIPANOVA;
proc genmod data=all;
  by Sample;
  class pop;
  model Y = pop / dist=zip type3;
  lsmeans pop / diff;
  zeromodel;
  *output out=ZIPReg pred=ZIPpred pzero=ZIPProb0;
  title2 'Generalized Linear Model Approach - ZIP Fit';
run;

```

```
%ODSON;
```

```

proc format;
  value sig 0 - .05 = 'Reject'
           .05 <- 1 = 'DNR';
run;

```

```

*****
*****      Rejection Rate for All Samples      *****;
*****

```

```

proc freq data=YANOVA scale;
  tables ProbF;
  format ProbF sig.;
  title2 'Rejection Rates for Treatment - ANOVA scale Y';

```

```

run;

proc freq data=PoissonANOVA;
  tables ProbChiSq;
  format ProbChiSq sig.;
  title2 'Rejection Rates for Treatment - Poisson w/ Scale=Pearson';
run;

proc freq data=ZIPANOVA;
  tables ProbChiSq;
  format ProbChiSq sig.;
  title2 'Rejection Rates for Treatment - ZIP Model';
run;

*****
*****                               Realized Effect Size                               *****
*****

proc transpose data=Univariate_all out=Univariate_all_T (drop=_LABEL_
rename=(_NAME_=Variable));
  title2 'Realized Effect Size';
run;

data Univariate_all_T;
  set Univariate_all_T;
  diff = COL1 - COL5;
run;

proc print data=Univariate_all_T;
  where Variable = 'mean';
run;

*****
*****                               Run Time                               *****
*****

%let t1 = %sysfunc(datetime());
%let elapsedTime = %sysevalf(&t1-&t0);
%put &elapsedTime;

```

Negative Binomial Samples

```
*****;
/*
Simulation Parameters-----

    pop = population ID
    p = percentage of zero values in population
    n = sample size for population
    mu = mean of normal distribution for population
    sigma = standard deviation of normal dist for population
    samples = number of samples to simulate

*/
*****;

options threads cpubcount=2;
options mprint symbolgen formchar="|----|+|----+=|-\<>*" ;

libname lyzn 'C:\Users\Lucas Young\Desktop\Thesis\Data';

%let t0 = %sysfunc(datetime()); *** Collect timing data. ***;

ods html close;
*ods html;
dm 'log; clear; output; clear;';
ods html path='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods html gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';
ods listing gpath='C:\Users\Lucas Young\Desktop\Thesis\Junk';

*ods trace on / listing;

%macro mix_NegBin(pop,p,n,ps,k,samples);
%global gp&pop gn&pop glambda&pop gsamples&pop;
%let gp&pop = %sysfunc(compress(&p,.));
%let gn&pop = &n;
%let glambda&pop = &lambda;
%let gsamples&pop = &samples;

data gen_data_&pop;
    pop=&pop;          *** Population ID          ***;
    p=&p;              *** Proportion of zeroes in pop
***;

```

```

n=&n;          *** Sample size from pop 1
***;
ps=&ps;        *** prob success for NegBin distribution
***;
k=&k;          *** k for NegBin distribution
***;
samples=&samples; *** Number of samples from pop
***;

call streaminit(0);

do sample=1 to &samples;
do i=1 to &n;
D=rand('bernoulli',&p); *** Use Bernoulli to select dist. ***;
if D=0 then Y=rand('negbinomial',&ps, &k);
else if D=1 then Y=0;
output;
end;
end;
run;

ods graphics on;

%mend mix_NegBin;

%macro ODSOff();
ods graphics off; ods exclude all; ods noresults;
options nonotes nosource nosource2 errors=0;
%mend;

%macro ODSOn();
ods graphics on; ods exclude none; ods results;
options notes source source2 errors=20;
%mend;

%mix_NegBin(1, 0.1, 3, 0.5, 32.6, 10000);
%mix_NegBin(2, 0.1, 3, 0.5, 32.6, 10000);
%mix_NegBin(3, 0.1, 3, 0.5, 32.6, 10000);
%mix_NegBin(4, 0.1, 3, 0.5, 32.6, 10000);

%mix_NegBin(5, 0.1, 3, 0.5, 20, 10000);
%mix_NegBin(6, 0.1, 3, 0.5, 20, 10000);
%mix_NegBin(7, 0.1, 3, 0.5, 20, 10000);
%mix_NegBin(8, 0.1, 3, 0.5, 20, 10000);

```



```

class pop;
model Y = pop / dist=normal type3 noscale scale=pearson;
lsmeans pop / diff;
title2 'Traditional ANOVA Approach';
run;

*****
*****      Poisson fit with Pearson Scaling      *****;
*****

ods output ModelANOVA = PoissonANOVA;
proc genmod data=all;
  by sample;
  class pop;
  model Y = pop / dist=poisson scale=pearson type3;
  lsmeans pop / diff;
  output out=PearPReg pred=PearPPred;
  title2 'Generalized Linear Model Approach - Scaled Poisson Fit';
run;

*****
*****      Zero-Inflated Poisson Fit      *****;
*****

ods output ModelANOVA = genmodZIP;
proc genmod data=all;
  by Sample;
  class pop;
  model Y = pop / dist=zip type3 scale=pearson;;
  lsmeans pop / diff;
  zeromodel;
  title2 'Generalized Linear Model Approach - ZIP Fit';
run;

%ODSON;

proc format;
  value sig 0 - .05 = 'Reject'
           .05 <- 1 = 'DNR';
run;

```

```
*****;
*****      Rejection Rate for All Samples      *****;
*****;
```

```
proc freq data=YANOVA scale;
  tables ProbF;
  format ProbF sig.;
  title2 'Rejection Rates for Treatment - ANOVA scale Y';
run;
```

```
proc freq data=PoissonANOVA;
  *where (Parameter='pop' & Level1='1');
  tables ProbChiSq;
  format ProbChiSq sig.;
  title2 'Rejection Rates for Treatment - Poisson w/ Scale=Pearson';
run;
```

```
proc freq data=genmodZIP;
  tables ProbChiSq;
  format ProbChiSq sig.;
  title2 'Rejection Rates for Treatment - genmodZIP Model';
run;
```

```
*****;
*****      Realized Effect Size      *****;
*****;
```

```
proc transpose data=Univariate_all out=Univariate_all_T (drop=_LABEL_
rename=(_NAME_=Variable));
  title2 'Realized Effect Size';
run;
```

```
data Univariate_all_T;
  set Univariate_all_T;
  diff = COL1 - COL5;
run;
```

```
proc print data=Univariate_all_T;
  where Variable = 'mean';
run;
```

```
*****;
*****      Run Time      *****;
```

```
%let t1 = %sysfunc(datetime());  
%let elapsedTime = %sysevalf(&t1-&t0);  
%put &elapsedTime;
```