Type I Error Assessment And Power Comparison Of Anova And Zero-Inflated Methods On Zero-Inflated Data

Lucas Young[#1], Rhonda Magel[#2], Curt Doetkott[#3]

[#1 2]*Department of Statistics* [#3]*Information Technology Services North Dakota State University Fargo, North Dakota 58108 United States*

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 otherwise might follow a normal or a Poisson distribution. 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 the Poisson zero-inflated model to a standard ANOVA model and also to a regular Poisson model on different types of zero-inflated data. Underlying distributions, number of populations, 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.*

Keywords - *Poisson, Normal, Zero-inflated Poisson model, ANOVA.*

I. 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' [1]. 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 could be misleading. Analysis of variance (ANOVA) will allow for small deviations from normal populations and still give accurate results, but the effect of large percentages of zeros on ANOVA 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 [2, 3, 4].

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 [5]. 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 they have not been compared to the results from using the standard ANOVA. Examples of zero-inflated models include hurdle, Tobit, and zero-inflated Poisson [6, 7, 8]. 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.

The purpose of this research is to compare the results from ANOVA with the results from the Poisson model, and the zero-inflated Poisson model (ZIP) when the underlying distributions are either normal or Poisson with various percentages of zero-inflation. 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 from both the normal and the Poisson distributions with varying levels of zero-inflation and sample sizes. For each combination of underlying distribution, level of zero-inflation, and sample size, we will estimate the Type

1 errors in the ANOVA and estimate the Type 1 errors using the regular Poisson model and using ZIP. Throughout the study, the stated Type 1 error is always 0.05. Thus, we will compare our estimated Type 1 errors with 0.05. For cases in which the Type 1 errors are approximately 0.05 or less for all the models, we will compare the powers. 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 the distribution 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.

Type 1 errors will be estimated for two and eight populations. Powers will be estimated for two populations for various changes in means, various sample sizes, and various zero-inflation percentages, when appropriate. In the case of eight populations, powers will be estimated for various sample sizes and zero-inflation percentages, where appropriate, when the first mean is different and all the remaining means are the same, and then when the first four means are equal, and the last four means are equal, but different from the first four.

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

1) How does Type I error compare between the ANOVA, the Poisson model, and the ZIP when the underlying distributions are normal or Poisson, but with various percentages of zero-inflated data?
2) Does the ZIP model, which is designed for zero-inflated Poisson data exhibit improved power over ANOVA or using a regular Poisson model across different percentages of zero-inflation for both normal and Poisson distributions?
3) What level of zero-inflation significantly affects the Type I error and power of ANOVA?

II. Background

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 defects in manufacturing, and healthcare expenditures are two of these areas [2, 3]. Regarding defects in manufacturing, when a machine is properly aligned the number of defects it will produces is consistently close to zero. When a machine is misaligned however, the number of defects it will produce may follow a Poisson distribution [3]. 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 [9, 10]. Finally, zero-inflated distributions can also be found in motor vehicle crash records [4]. 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. 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 [11].

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 [6]. A further description of hurdle models can be found in Mullahy [12].

Another approach to dealing with zero-inflation is the Tobit model [6] which originated for application with truncated data [13]. 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 [14] it can be applied to zero-inflated data [15].

The ‘zero-inflated Poisson’ (also referred to as ‘ZIP’) model has been proposed as a solution for dealing with zero-inflation in a Poisson distribution [7]. With reference to manufacturing machine defects, the zero-inflated Poisson model 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 [3]. Greene also discusses the importance of distinguishing between zero-inflated and over dispersed distributions

[7]. A method for dealing with over dispersed Poisson or lack of fit to a Poisson distribution is to make likelihood adjustments [11]. Though papers discussing the comparison of zero-inflated models exist [1, 2], a simulation study to quantify the relative Type I error and power while considering the variables of interest will provide additional insight.

III. Simulation Study

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 the ANOVA, Poisson model and ZIP model 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 two types of underlying distributions with varying levels of zero-inflation when simulating the samples. The distributions we will consider are normal and Poisson. 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 (μ_2, μ_3, μ_4, μ_5, μ_6, μ_7, and μ_8) and the remaining 7 means are equal; this will be referred to as the '1 vs 7' scenario. The second scenario is when the first four means (μ_1, μ_2, μ_3, and μ_4) are equal, and the last four means (μ_5, μ_6, μ_7, and μ_8) are equal but different than the first four; 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 that are different than the control, while 3 treatment populations do not differ from the control.

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 (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 [6].

Simulation Parameters

To select the underlying parameters for the two 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. Though, the variance is not the same between the two distributions, the mean is 20 for each.

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: At least one μ_x 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 = 0.5σ). 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 = 1σ. 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 the 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: At least one μ_x 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
   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
   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
   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
   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
   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

IV. 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. 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

Eight-Population Type I Error Assessment

The results of the eight-population Type I error assessment on normal samples can be found in Tables 8, 9, and 10. 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 8. 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 9. 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 10. 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 11, 12, and 13. In Table 12 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 11. 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 12. 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 13. 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

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 1, 2, and 3 for percentage of zero inflation being 0%, 20%, and 50%. The relative power of the models at the 1σ effect size were amplified in the 2σ effect size simulations.

For eight-population zero-inflated normal samples, we only saw acceptable Type I error for samples of n = 10 when using the Poisson model. Subsequently, we will only compare models for this distribution at the n = 10 sample size. The zip model had low power in comparison to the other models when the zero-inflation percentage was zero. There is no distinct power advantage for any of the models for higher percentages of zero-inflation. Figures 4 and 5 are given for the powers for each model for 20% and 50% zero-inflation. 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).

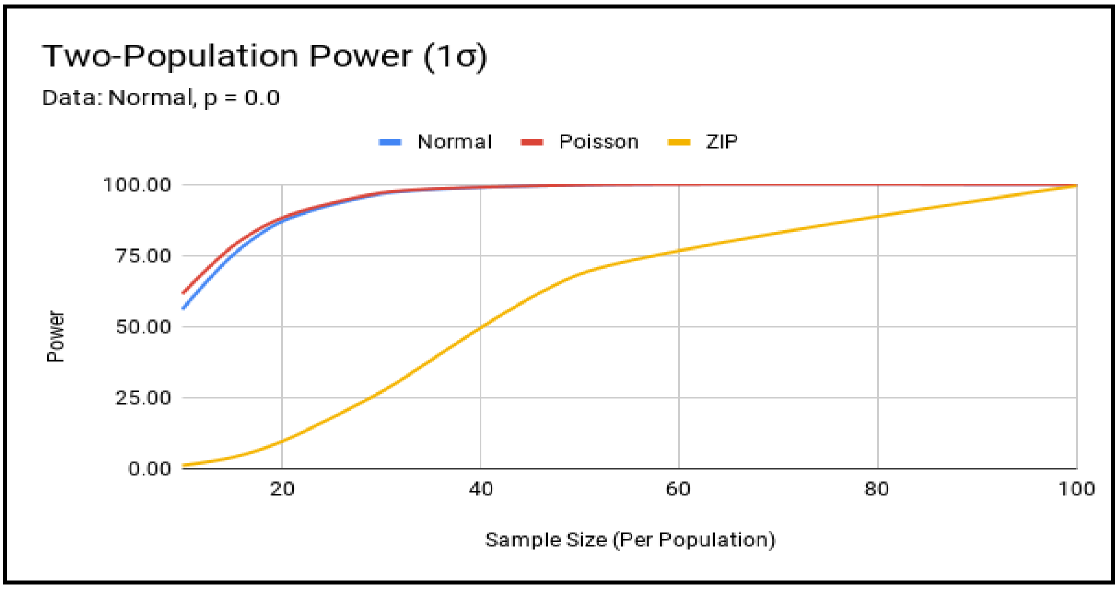

Figure 1. Normal Sample – p = 0.0 Two-Population Power (1σ).

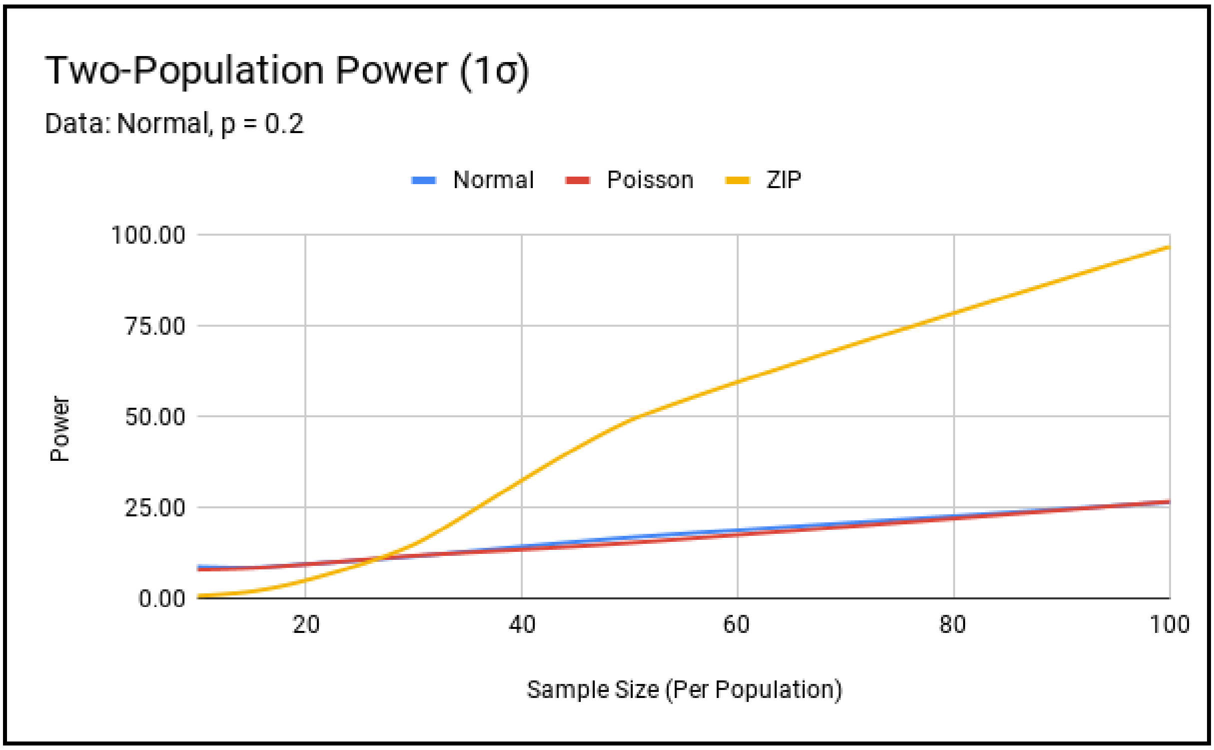

Figure 2. Normal Sample – p = 0.2 Two-Population Power (1σ).

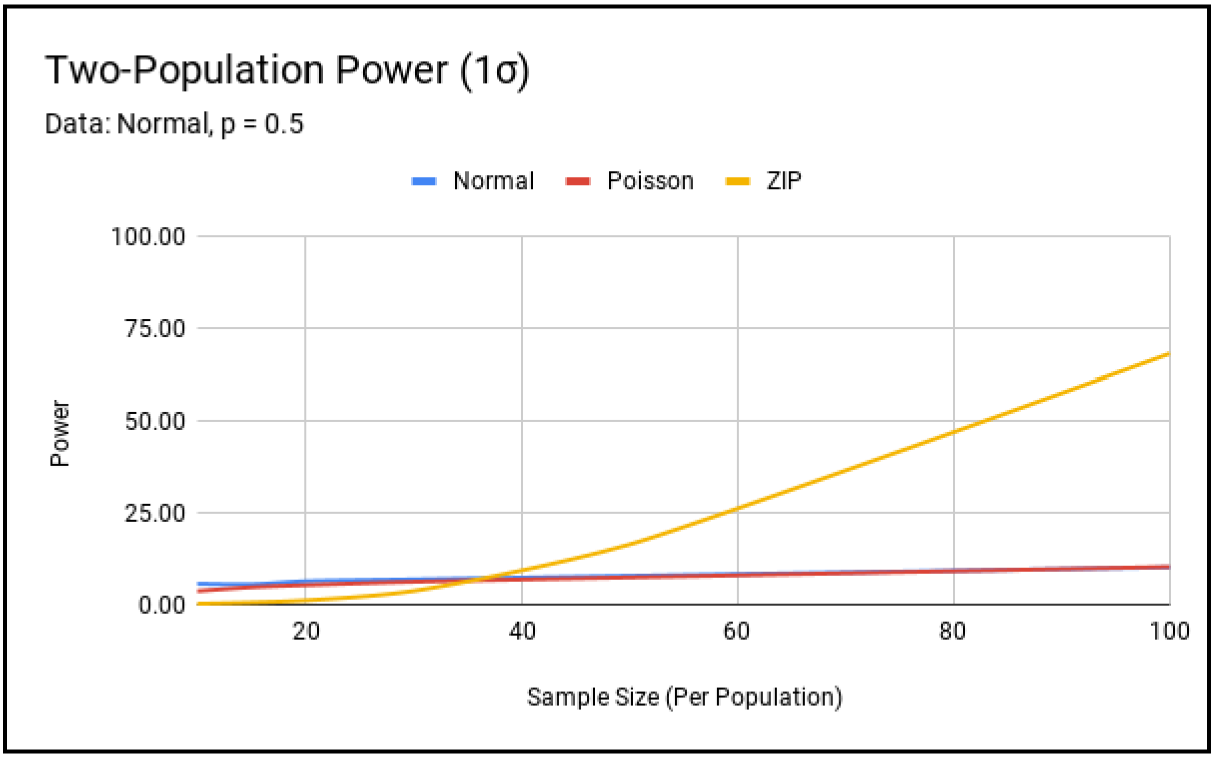

Figure 3. Normal Sample – p = 0.5 Two-Population Power (1σ).

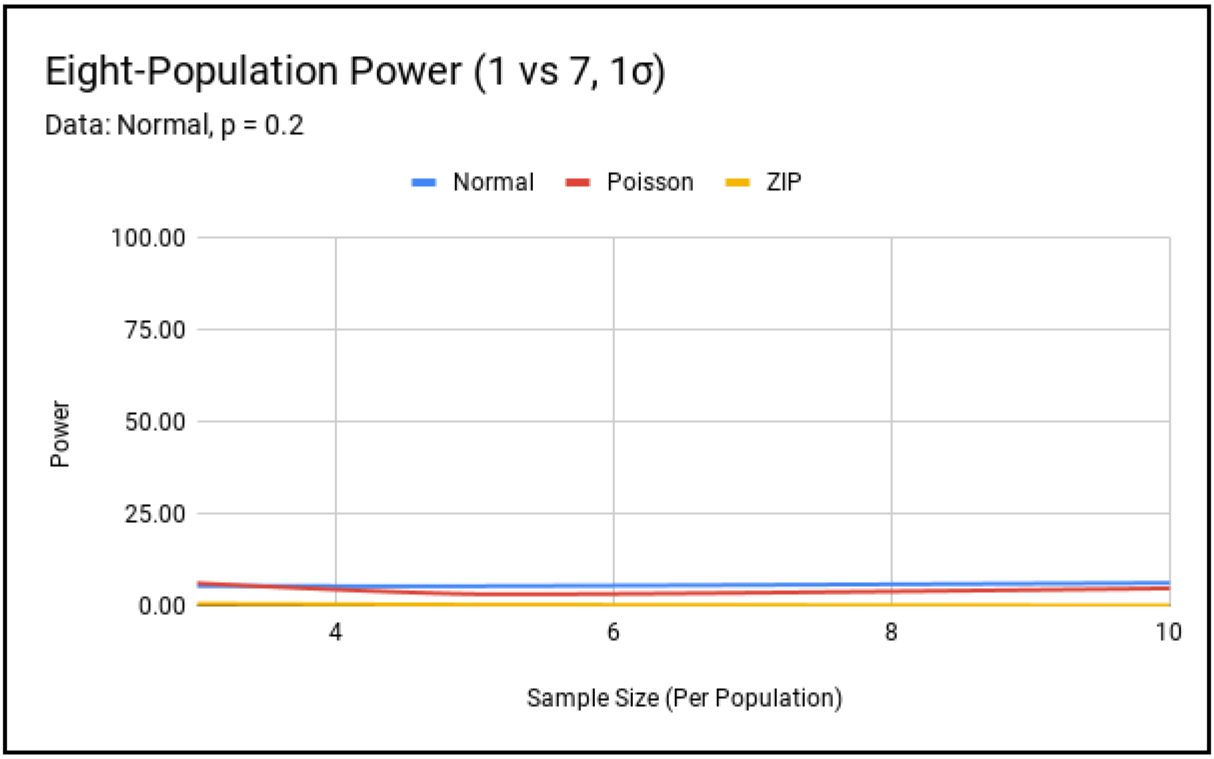

Figure 4. Normal Sample – p = 0.2 Eight-Population Power (1 vs 7, 1σ).

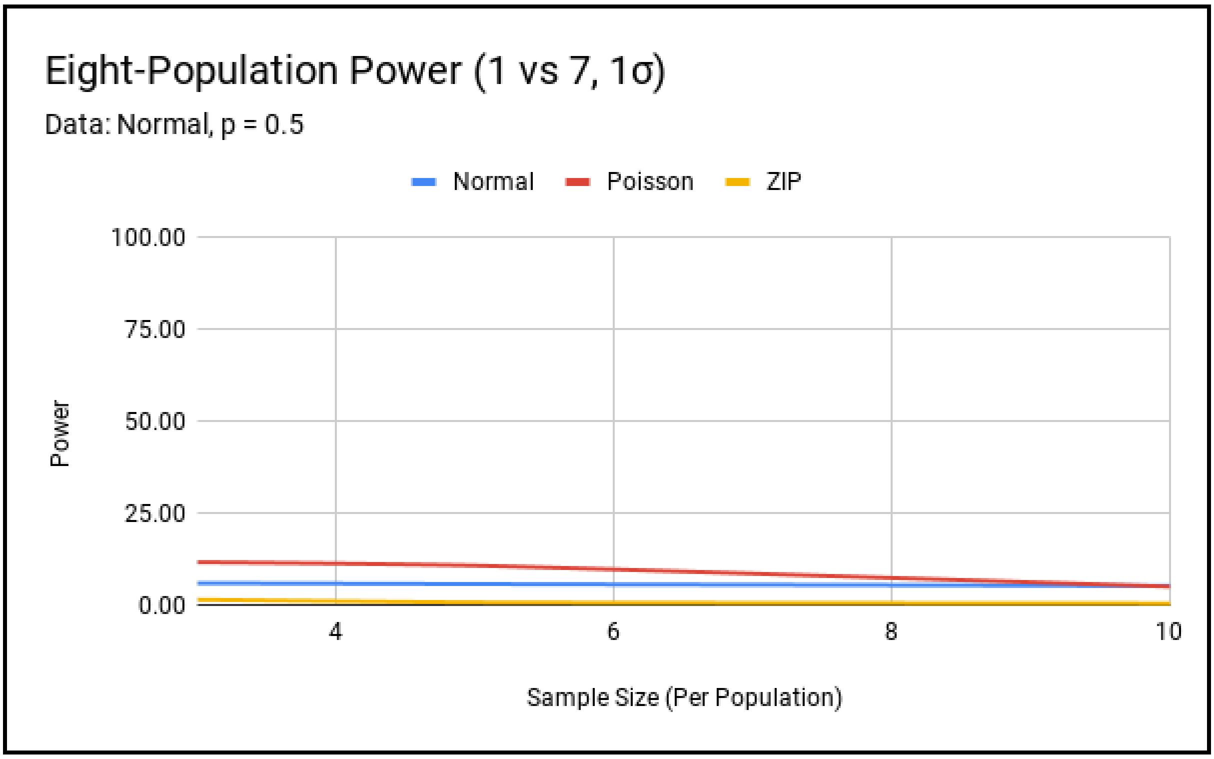

Figure 5. Normal Sample – p = 0.5 Eight-Population Power (1 vs 7, 1σ).

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 (except at 0%) and sample sizes. . Figures

6 and 7 show the comparison the powers of all three models for the 1σ effect size when the percentage of zero-inflation was 20% and 50%.

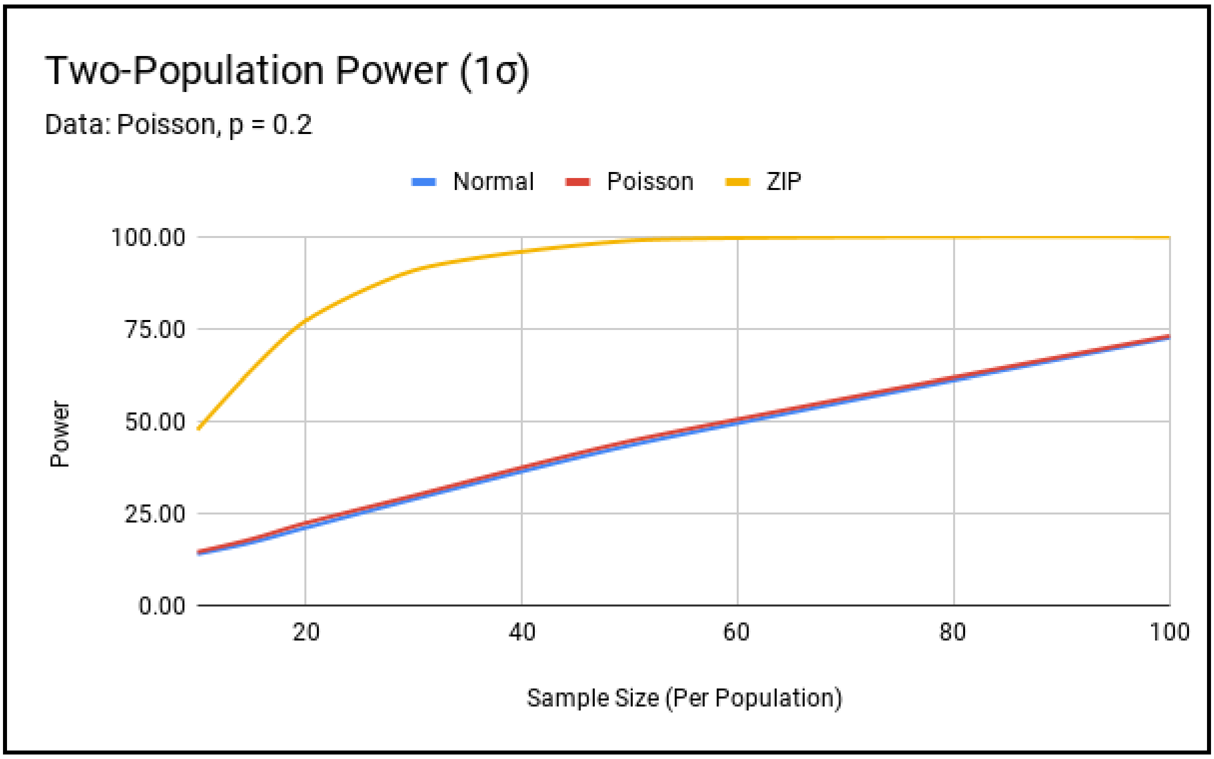

Figure 6. Poisson Sample – p = 0.2 Two-Population Power (1σ).

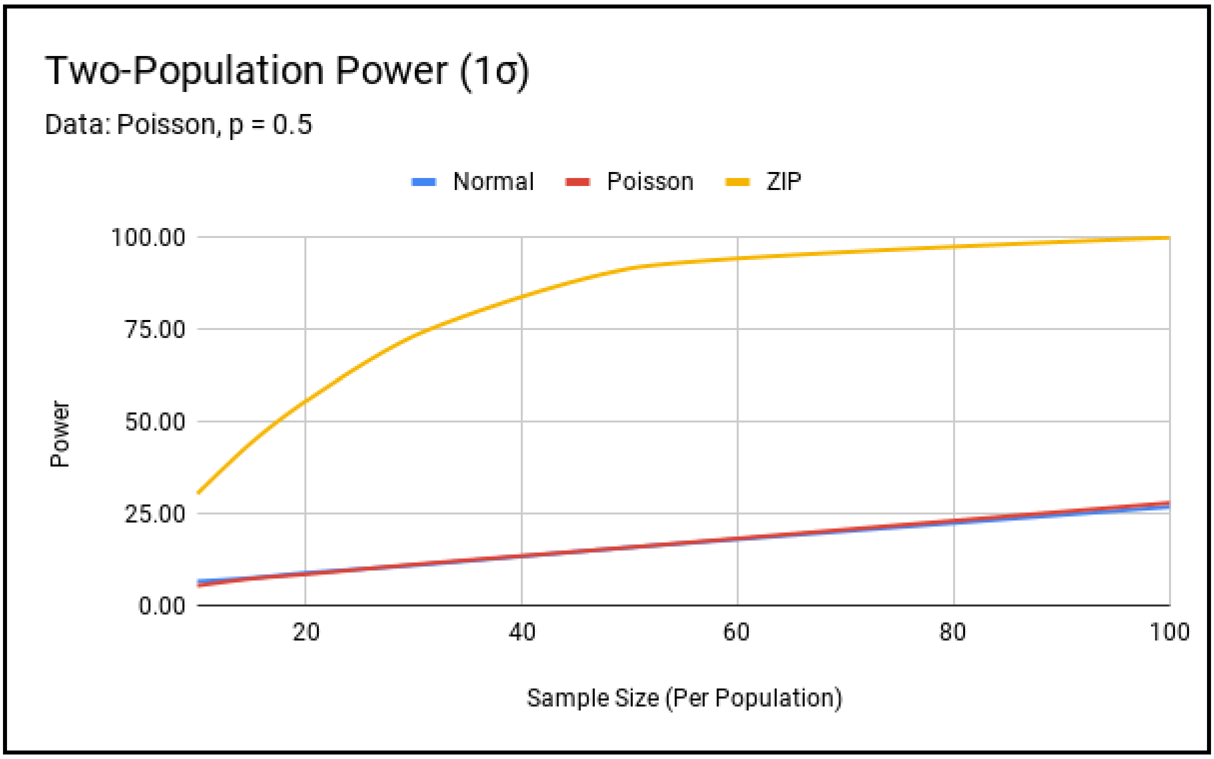

Figure 7. Poisson Sample – p = 0.5 Two-Population Power (1σ).

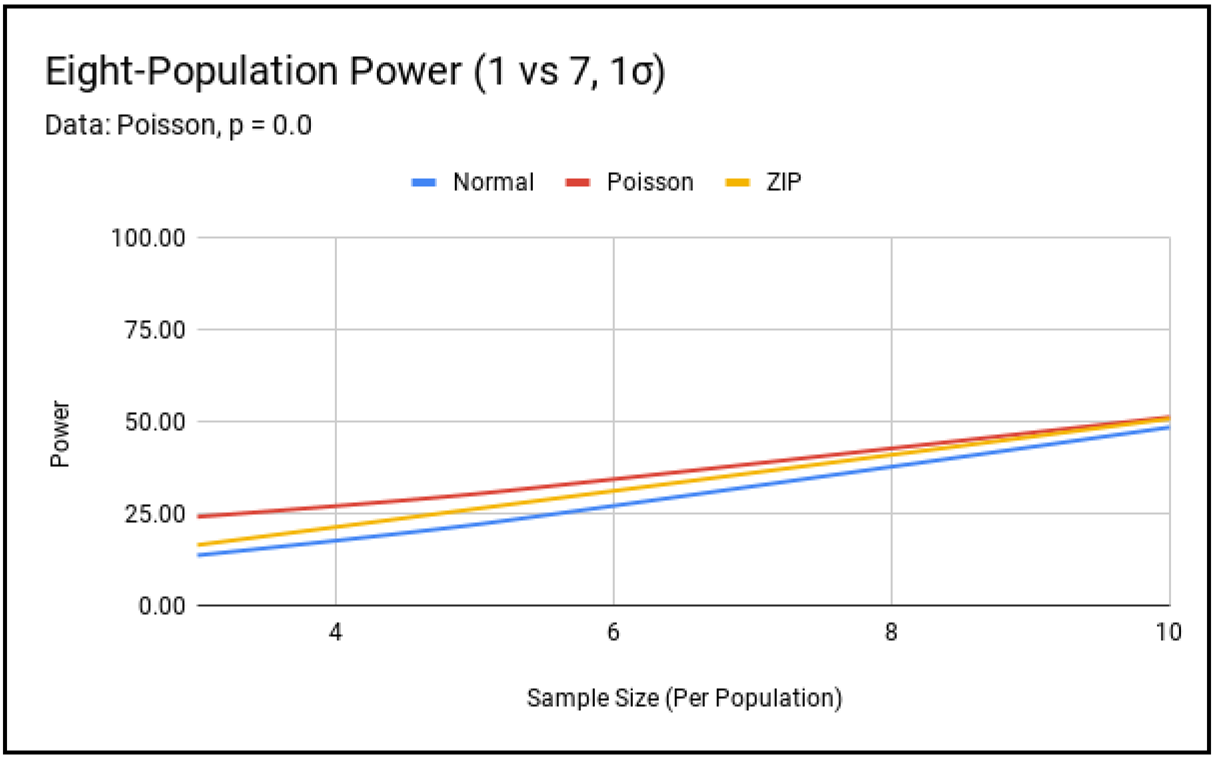

Figure 8. Poisson Sample – p = 0.0 Eight-Population Power (1 vs 7, 1σ).

For eight-population zero-inflated Poisson samples, we only saw acceptable Type I error for samples of n = 10 for the Poisson model. 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 8, 9, and 10 for percentages of zero-inflation of 0%, 30% and 50%.

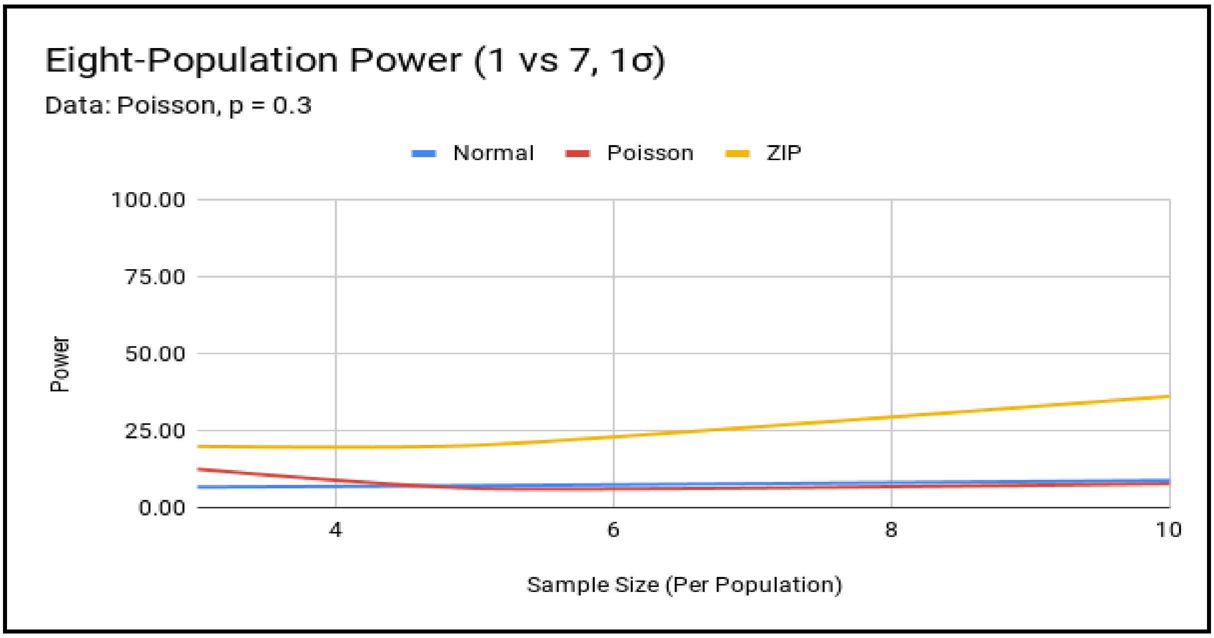

Figure 9. Poisson Sample – p = 0.3 Eight-Population Power (1 vs 7, 1σ).

V. Conclusions

Research Questions

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

1) How does Type I error compare between the ANOVA, the Poisson model, and the ZIP when the underlying distributions are normal or Poisson, but with various percentages of zero-inflated data?
2) Does the ZIP model, which is designed for zero-inflated Poisson data, exhibit improved power over ANOVA or using a regular Poisson model across different percentages of zero-inflation for both normal and Poisson distributions?
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 sizes 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 the normal model. Type 1 error was controlled for the Poisson model and the ZIP model for sample sizes of at least 10. To more specifically answer the Type I error research question, for the models and two 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.

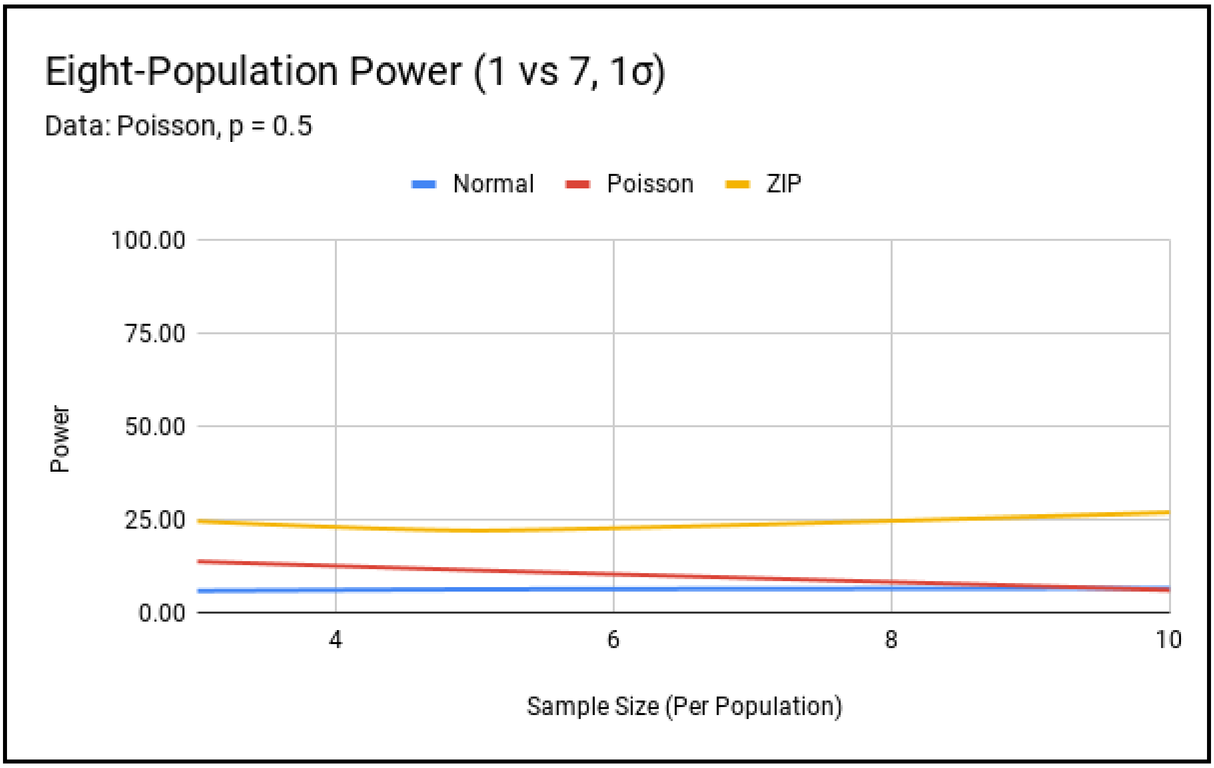

Figure 10. Poisson Sample – p = 0.5 Eight-Population Power (1 vs 7, 1σ).

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 two sample distributions. 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.

REFERENCES

[1] Min, Y. & Agresti, A. (2002) Modeling Nonnegative Data with Clumping at Zero: A Survey. *Journal of Iranian Statistical Society* Vol. 1, pp. 7-33.

[2] Eggers, J. (2015). On Statistical Methods for Zero-Inflated Models. Technical Report U.U.D.M. Project Report 2015:9, Uppsala University.

[3] Lambert, D. (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* Vol. 34, No. 1, pp. 1-14.

[4] 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.

[5] Fisher, R. (1963) Statistical Methods for Research Workers, 13th ed. Edinburgh: Oliver and Boyd.

[6] Stroup, W. (2013) Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. CRC Press.

[7] Tobin, J. (1958) Estimation of Relationships for Limited Dependent Variables. *Econometrica* Vol. 26, No. 1, pp. 24-36.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Mullahy, J. (1986) Specification of Some Modified Count Data Models. *Journal of Econometrics* Vol. 33, pp. 341-365

[13] Amemiya, T. (1973) Regression Analysis when the Dependent Variable Is Truncated Normal. *Econometrica* Vol. 41, No. 6, pp. 997-1016.

[14] McDonald, F. & Moffitt, R. (1980) The Uses of Tobit Analysis. *The Review of Economics and Statistics* Vol. 62, No. 2, pp. 318-321.

[15] Hall, D. & Zhang, Z (2004) Marginal models for zero inflated clustered data. *Statistical Modeling* Vol. 4, pp. 161-180