APPENDIX III.C.1 - FILTER ADOPTION

DRAFT FOR PUBLIC COMMENTS

Version 1.0: July 11, 2019

DENVER WATER

Appendix III.C.1

DRAFT 4- Lead Filter Program Sample Size Required for Determining Rate

Date:	Revised July 3, 2019 March 14, 2019
To:	Nicole Poncelet-Johnson Alexis Woodrow
From:	Chad Seidel Chris Corwin Carleigh Samson

Executive Summary

The objective of this memorandum is to develop a statistical method to estimate the number of Denver Water customers that adopt a lead filter and therefore reduce their exposure to lead in their drinking water. To meet this objective, the memorandum answers the following question: "How many Denver Water customers must respond to the lead filter program survey to sufficiently estimate filter adoption rate for all customers provided a lead filter?" Survey responses from 1,059 or more randomly selected Denver Water customers that received lead filters are needed to estimate the filter adoption rate (p) with at least 95% confidence and no more than 5% error, based on an adoption rate greater than 60%. Distributing the survey to a group of 1,250 Denver Water customers that received a lead filter is recommended to achieve the requisite survey response from 1,059 random surveyed customers while limiting the self-selection bias.

Introduction

Corona prepared a statistical approach to support Denver Water's efforts in understanding the required number of customers to be surveyed to sufficiently estimate point of use (POU) device adoption rates. Denver Water is investigating a program to provide POU devices to customers to protect them from lead exposure. When used properly, POU devices are effective at removing dissolved and particulate lead from drinking water. Therefore, the effectiveness of the POU devices in protecting Denver Water customers from lead exposure relies on customers' adoption of the devices. POU device adoption assumes customers are installing, using and maintaining the device properly, as well as replacing the filters at the appropriate time. Customers not using the POU but relying on bottled water for drinking and cooking will also be considered an adoption. Corona's statistical approach described in this memorandum details the number of customers that received a POU device that need to be surveyed based on the acceptable confidence level and the error in the estimated POU device adoption rate.

Statistical Analysis

The objective of the statistical analysis is to estimate the number of Denver Water customers that adopt their lead filter and therefore reduce their exposure to lead in their drinking water. The total number of Denver Water customers that adopt their lead filter can be estimated using the total number of Denver

Water customers that receive a lead filter and the filter adoption rate for this entire population. To avoid having to survey the entire population of customers receiving a lead filter, a statistical analysis can be used to estimate the filter adoption rate utilizing a subset of the population. To determine the subset sample size required, the adoption rate distribution, confidence level, and acceptable error must be considered.

Lead filter adoption takes on a binomial distribution of "adoption" or "lack of adoption" (e.g. 1 or 0), which gives a discrete probability distribution of Bernoulli trials. A Bernoulli trial is an event that has two possible outcomes, such as flipping a coin. Each filter adoption, or lack of adoption, can be described as a Bernoulli trial because there are only two possible outcomes: adoption ("success") or no adoption ("failure"). We assume each customer's lead filter adoption, or lack of adoption, is independent of other customers' filter device adoption, and therefore, each "trial" constitutes a random, independent experiment. This assumption that each customer's filter adoption is not dependent on any other's customer's filter adoption emphasizes the need for Denver Water to ensure that surveyed customers are randomly selected. More information on the recommended survey procedure to prevent sampling bias is provided in the following section. We also assume that the probability, p, of a success in each trial remains constant. This means that we assume the probability that each customer will adopt the lead filter is constant and equal to some value p. Because actual adoption may not be constant, we recommend Denver Water repeat the survey annually.

The binomial distribution has a mean np and variance $np(1-p)$ where n is number of Bernoulli trials. In the context of this memorandum, n is equal to the number of surveyed Denver Water customers offered lead filters. The number of surveyed customers who have adopted their filter are defined as X, where $X \leq n$. The quantity X/n is the point estimator ($\hat{P}$) of the filter adoption rate (p) for all customers receiving a lead filter. The binomial distribution is described in further detail in the Appendix. The descriptions were developed utilizing Montgomery & Runger (2007)[1].

The size of the confidence interval, which can also be defined as the difference between the true proportion of all Denver Water customers' filter adoption rate, p, and the proportion of surveyed customers' lead filter adoption rate, $\hat{p}$, is dependent on both α, which defines the confidence level, and n, the sample size of surveyed customers[2]. By defining the error $E = \left|p - \hat{P}\right|$ and selecting an acceptable error (i.e. 0.05) and an acceptable statistical power (i.e. 95%) that E is less than our acceptable error, we can determine the required sample size utilizing the statistical computing software R package Binomial Confidence Intervals For Several Parameterizations ("binom")[3]. The power of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true, which can be interpreted as the probability of correctly rejecting the false null hypothesis. In this application, the alternative hypothesis is true if the true proportion of all Denver Water customers' adoption rate p is greater than the proportion of surveyed customers' adoption rate $\hat{p}$ minus the error E.

$$H_0: p = \hat{p} - E$$

---

[1] Montgomery, D.C. & Runger, G.C. 2007. Applied Statistics and Probability for Engineers: Fourth Edition. John Wiley & Sons, Inc. USA.

[2] Note that $\hat{P}$is a random variable point estimator for the POU device adoption rate (p) and $\hat{p}$ is the POU device adoption rate for surveyed Denver Water customers.

[3] Dorai-Raj, S. 2015. Package 'binom'. "Binomial Confidence Intervals For Several Parameterizations. Accessed 5/8/2019. https://cran.r-project.org/web/packages/binom/binom.pdf

$$H_0 : p > \hat{p} - E$$

Figure 1 illustrates the relationship between the survey sample size of Denver Water customers needed to estimate the filter acceptance rate and the acceptable error and confidence level. Table 1 summarizes the required sample sizes for acceptable errors of 10%, 5%, and 1% and for filter adoption rates of 50%, 60%, 70%, 80%, 90%, and 95% assuming 95% statistical power. To ensure, with 95% statistical power, that the sample size filter device acceptance rate, of at least 60%, is within 5% of the entire customer population filter acceptance rate, responses would be required from 1,059 random surveyed customers. If the adoption rate falls to 60% with the sample size of 1,059 random surveyed customers, the error of the estimate increases by less than 1%. Therefore, we recommend a sample size of random surveyed customers of 1,059 to be both achievable and representative.

Figure 1 Survey sample size based on acceptable error and filter adoption rate using a binomial distribution assumption with 95% statistical power

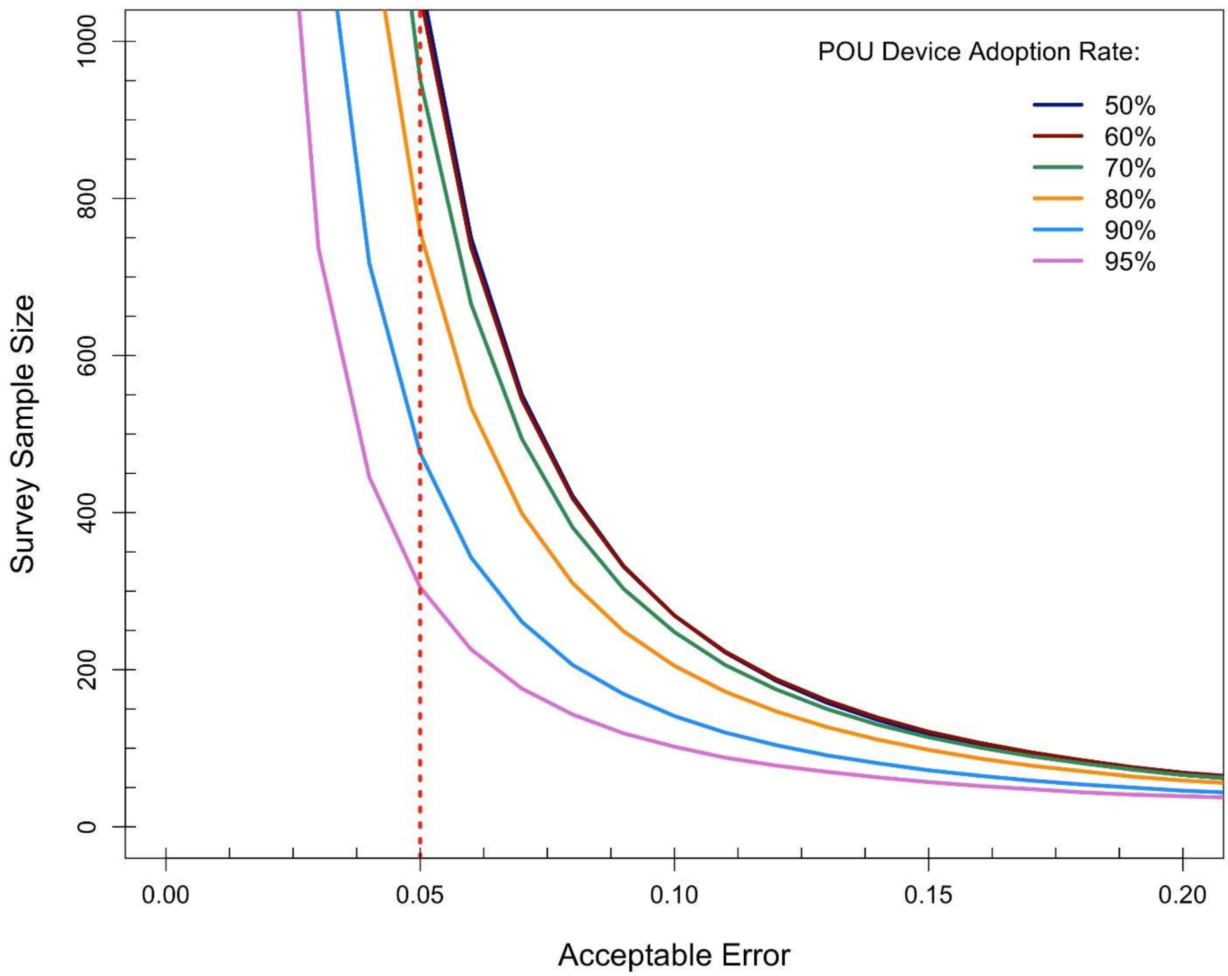


Table 1 Required sample size of surveyed customers based on acceptable error and filter adoption rate and 95% statistical power

Acceptable Error	50% Adoption	60% Adoption	70% Adoption	80% Adoption	90% Adoption	95% Adoption
10%	n=269	n=269	n=248	n=205	n =141	n=102
5%	n=1,081	n=1,059	n=951	n=757	n=476	n=305
1%	n=27,054	n=26,080	n=22,942	n=17,640	n=10,173	n=5,630

Alternatively, a normal approximation can be assumed for the point estimator $\hat{P}$ of the filter adoption rate (p) for all customers receiving a lead filter if the sample size (n) is sufficiently large and p is not too close to 0 or 1. To apply this approximation, we require that np and $n(1-p)$ are greater than or equal to 5. The normal distribution, standard normal distribution and the normal approximation and confidence interval for the probability p that each customer will adopt the filter are described in detail in the Appendix.

Using the normal approximation, the size of the confidence interval, which can also be defined as the difference between the true proportion of all Denver Water customers' filter adoption rate, p, and the proportion of surveyed customers' filter adoption rate, $\hat{p}$, is dependent on both α, which defines the confidence level, and n, the sample size of surveyed customers[4]. If we define the error $E = \left|p - \hat{P}\right|$, and we select an acceptable error (i.e. 0.05) and an acceptable confidence (95%) that *E* is less than our acceptable error, we can determine the required sample size as:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

Equation 1

Using the exact binomial distribution results in a more conservative sample size requirement as compared with the normal approximation assumption. Therefore, if the surveyed customers' filter adoption rate is greater than 60% a sample size of 1,059 for survey responses from Denver Water customers is a sufficiently conservative requirement to determine that the filter adoption rate for all Denver Water customers receiving a lead filter is within 5% of the surveyed customers' filter adoption rate.

Sample Selection and Verification

A random selection of 1,250 customers from the group of all of the customers provided a filter should be performed each year. The customers selected and the corresponding surveys received should be randomized based on geography and demographics. Efforts to achieve the requisite response rates (e.g. at least 1,059 of 1,250) must be undertaken to prevent self-selection bias in the reporting group. These efforts may include mailings, phone calls, and site visits to the randomly selected customers. Community groups present an opportunity to leverage independent parties that might obtain higher response rates and a higher level of truthfulness in the responses.

Even though the number of respondents may approach a level of confidence and error that are acceptable, efforts should be continued to complete responses from all the customers selected for

[4] Note that $\hat{P}$is a random variable point estimator for the POU device adoption rate (p) and $\hat{p}$ is the filter adoption rate for surveyed Denver Water customers.

verification. A high response rate from the random selection ensures a full representation of the diversity of Denver Water's customer base.

Recommendation

For the lead filter program to be considered successful, the adoption rate needs to be greater than or equal to 60% for equivalence. However, Denver Water should continue efforts to maximize the adoption rate. Corona recommends obtaining responses from a minimum of 1,059 customers out of a randomly selected group of 1,250. The survey should be repeated on an annual basis to detect changes in adoption rate over time. Responses from 1,059 randomly selected customers would achieve 95% confidence that the true sample adoption is within 5% of the subsample adoption if the subsample adoption is above 60%. If the subsample adoption is greater than 60%, then the confidence is increased and/or the error is decreased. If the subsample adoption is lower, then Denver Water should take measures to increase the adoption rate.

Appendix

Binomial Distribution

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent
(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"
(3) The probability of a success in each trial, denoted as p, remains constant

The random variable X that equals the number of trials that result in a success has a binomial random variable with parameters $0 < p < 1$ and n = 1,2,…. The probability mass function X is:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \textit{Equation 2}$$

$\binom{n}{x}$ equals the total number of different sequences of trials that contain x successes and $n - x$ failures. The total number of different sequences that contain x successes and $n - x$ failures times the probability of each sequence equals $P(X = x)$.

If X is a binomial random variable with parameters p and n, the mean $\mu = E(X) = np$ and the variance $\sigma^2 = V(X) = np(1-p)$.

Normal and Standard Normal Distributions

A normal random variable X from a normal distribution with mean μ and variance σ^2 can be standardized by the following:

$$Z = \frac{\bar{X} - \mu}{\sigma} \qquad \textit{Equation 3}$$

Z is then a standard normal random variable with a standard normal distribution. A standard normal distribution is a normal distribution with mean μ=0 and variance σ^2=1. The standard normal distribution probability density function is described below by $P(x)$:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2)/(2\sigma^2)} = \frac{1}{\sqrt{2\pi}} e^{\left(-x^2/2\right)} \qquad \textit{Equation 4}$$

To determine the probability that the standard normal random variable Z is less than or equal to some value z, written as $P(Z \leq z)$, we can use the cumulative distribution function of a standard normal random variable, denoted as $\Phi(z)$, which is found by integrating the probability density function:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\left(-z^2/2\right)} dz \qquad \textit{Equation 5}$$

To determine the probability that Z is greater than some value z, $P(Z > z)$, we can utilize the fact that the integral of the probability density function taken from -∞ to ∞ is equal to 1:

$$P(Z > z) = 1 - P(Z \leq z) \qquad \textit{Equation 6}$$

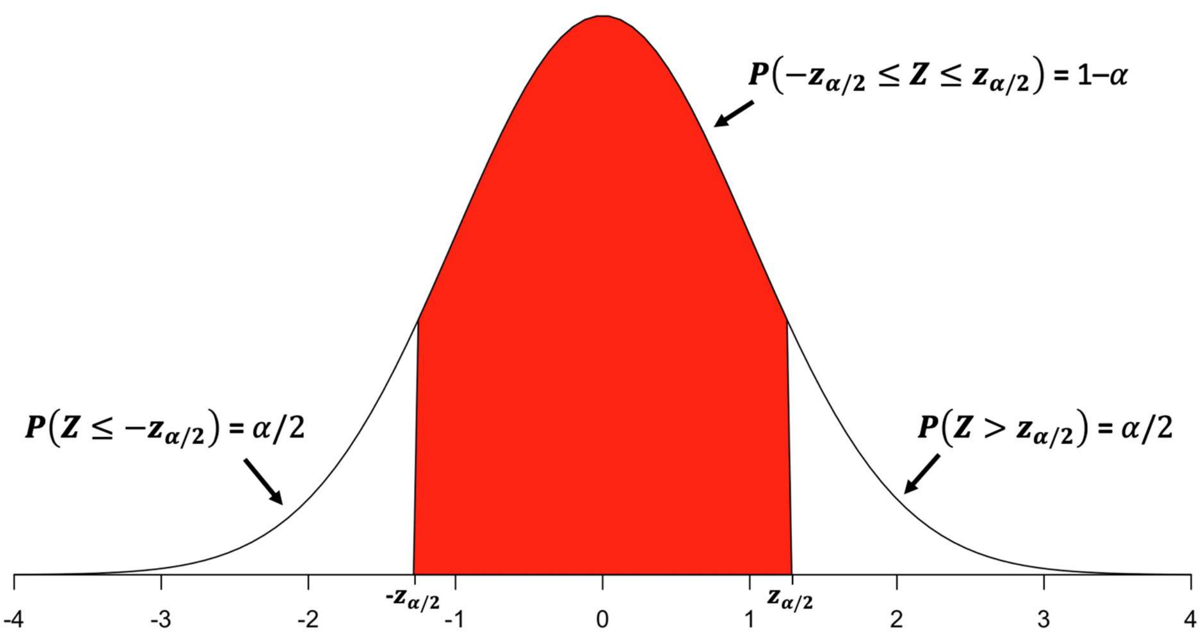
For a standard normal distribution, we can define the probability that Z is within a defined confidence interval with a confidence level of 100(1-α)% by:

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) \cong 1 - \alpha \qquad \textit{Equation 7}$$

where $z_{\alpha/2}$ is defined as the z value that corresponds with the upper $\alpha/2$ percentage point of the standard normal distribution (see Figure 2). Alternatively, we can say with 100(1-α)% confidence that:

$$-z_{\alpha/2} \leq Z \leq z_{\alpha/2} \qquad \textit{Equation 8}$$

Figure 2 Standard normal distribution showing confidence intervals for Z



Large Sample Confidence Intervals for the Mean of a Normal Distribution

In the case of sampling from a normally distributed population with an unknown mean and a known standard deviation σ with the objective to estimate the population mean, a large sample confidence interval for the mean μ can be determined if the sample size is sufficiently large. Given the assumption that the sample size is large, the central limit theorem can be applied such that the sample mean $\bar{X}$ has an approximate normal distribution with mean μ and variance σ^2/n. Therefore, for a normal distribution with a large sample size:

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \qquad \textit{Equation 9}$$

where: $\bar{X}$ is the sample mean,

μ is the distribution mean,

S is the sample standard deviation, and

n is the sample size.

The large sample confidence interval for μ for a confidence level of approximately 100(1-α)% can then be described as:

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq x \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$ *Equation 10*

Normal Approximation to the Binomial Proportion

Using the normal approximation, the sampling distribution of $\hat{P}$ is approximately normal with mean *p* and variance *p*(1-*p*)/n. If Denver Water uses a sufficiently large sample size n and *p* is not too close to 0 or 1, the normal approximation for *p*, the probability that each Denver Water customer who receives a lead filter will adopt the filter, is equal to the following:

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ *Equation 11*

where Z has an approximate standard normal distribution. Using the normal approximation of the binomial proportion, we can use the standard normal confidence intervals to determine the following approximate confidence interval for our binomial proportion, p:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ *Equation 12*

The error E between the true filter adoption rate among all Denver Water filter recipients p and the filter adoption rate among all surveyed Denver Water customers $\hat{p}$ can be defined as $E = \left|p - \hat{P}\right|$ where $\hat{P}$ is a random variable from a binomial distribution with mean *p* and variance *p*(1-*p*)/n. Thus, there is 100(1-α)% confidence that $E < z_{\alpha/2}\sqrt{p(1-p)/n}$. If we set $E = z_{\alpha/2}\sqrt{p(1-p)/n}$, we can solve for the sample size n.