

Water Wells

Zach White

STA 723 Case Studies

January 24, 2017

Introduction

- ▶ In many countries, people get their water from wells. Sometimes these wells are contaminated with arsenic, which when consumed in high concentrations causes skin and bladder cancer, as well as cardiovascular disease.

Introduction

- ▶ In many countries, people get their water from wells. Sometimes these wells are contaminated with arsenic, which when consumed in high concentrations causes skin and bladder cancer, as well as cardiovascular disease.
- ▶ In some cases, people living near contaminated wells have the opportunity to get water from nearby uncontaminated wells.

Introduction

- ▶ In many countries, people get their water from wells. Sometimes these wells are contaminated with arsenic, which when consumed in high concentrations causes skin and bladder cancer, as well as cardiovascular disease.
- ▶ In some cases, people living near contaminated wells have the opportunity to get water from nearby uncontaminated wells.
- ▶ We seek to design a study to assess how distance influences whether or not a household will switch.

Introduction

- ▶ In many countries, people get their water from wells. Sometimes these wells are contaminated with arsenic, which when consumed in high concentrations causes skin and bladder cancer, as well as cardiovascular disease.
- ▶ In some cases, people living near contaminated wells have the opportunity to get water from nearby uncontaminated wells.
- ▶ We seek to design a study to assess how distance influences whether or not a household will switch.
- ▶ We will assume the data from the previous class are from a previous, available study.

Power Analysis

- ▶ Power ($1 - \beta$) is the probability of rejecting a false null hypothesis
 - ▶ β is the probability of failing to reject a false null hypothesis
- ▶ Through power analysis, we want to understand how many participants are required to achieve desired levels of power.

Power Analysis

- ▶ Power ($1 - \beta$) is the probability of rejecting a false null hypothesis
 - ▶ β is the probability of failing to reject a false null hypothesis
- ▶ Through power analysis, we want to understand how many participants are required to achieve desired levels of power.
- ▶ We need the following
 - ▶ a designated α level, which is the probability of rejecting a true null hypothesis

Power Analysis

- ▶ Power ($1 - \beta$) is the probability of rejecting a false null hypothesis
 - ▶ β is the probability of failing to reject a false null hypothesis
- ▶ Through power analysis, we want to understand how many participants are required to achieve desired levels of power.
- ▶ We need the following
 - ▶ a designated α level, which is the probability of rejecting a true null hypothesis
 - ▶ a desired power level

Power Analysis

- ▶ Power ($1 - \beta$) is the probability of rejecting a false null hypothesis
 - ▶ β is the probability of failing to reject a false null hypothesis
- ▶ Through power analysis, we want to understand how many participants are required to achieve desired levels of power.
- ▶ We need the following
 - ▶ a designated α level, which is the probability of rejecting a true null hypothesis
 - ▶ a desired power level
 - ▶ some understanding of the targeted effect size

Power Analysis

- ▶ Power ($1 - \beta$) is the probability of rejecting a false null hypothesis
 - ▶ β is the probability of failing to reject a false null hypothesis
- ▶ Through power analysis, we want to understand how many participants are required to achieve desired levels of power.
- ▶ We need the following
 - ▶ a designated α level, which is the probability of rejecting a true null hypothesis
 - ▶ a desired power level
 - ▶ some understanding of the targeted effect size

Simulation

- ▶ We will calculate power using a simulation study from previous data.

Simulation

- ▶ We will calculate power using a simulation study from previous data.
- ▶ From the past data, we have for each individual i , $switch_i$, $arsenic_i$, $distance_i$, $education_i$, and $association_i$.

Simulation

- ▶ We will calculate power using a simulation study from previous data.
- ▶ From the past data, we have for each individual i , $switch_i$, $arsenic_i$, $distance_i$, $education_i$, and $association_i$.
- ▶ We will use sampling with replacement from this past data set to calculate power under varying conditions.

Simulation Scheme

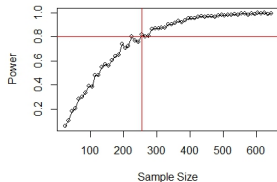
Estimate parameters using MLE from logistic model

Fix parameters with MLE or effect size

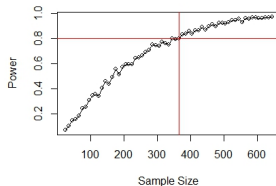
- ▶ Draw n data from previous data
- ▶ Calculate $\log(\frac{p_i}{1-p_i})$ given the drawn data and our fixed model
- ▶ calculate p_i
- ▶ Draw $switch_i \sim \text{bernoulli}(p_i)$
- ▶ Fit logistic model to fitted data
- ▶ Calculate p-value for $\beta_{distance}$
- ▶ Repeat R times
- ▶ Increase n
- ▶ Repeat

Results

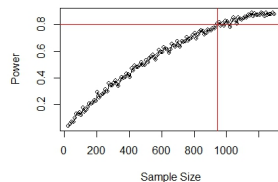
Effect size = -0.20



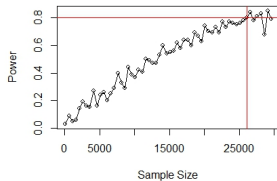
Effect size = -0.17



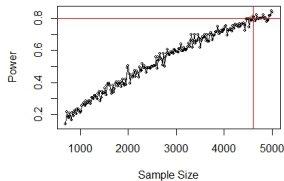
Effect size = -0.10



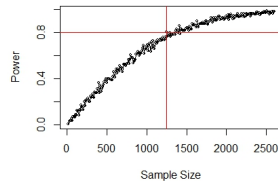
Effect size = -0.02



Effect size = 0.05



Effect size = 0.10



Conclusions and Limitations

Limitations

- ▶ We are assuming a logistic model when simulating the data.
- ▶ We may not have actual previous data
 - ▶ Alternative: Estimate covariance matrix and draw from multivariate normal

Conclusions

- ▶ Required sample size doesn't seem to be symmetric about effect size of zero.

Sample size required to
get power = 0.80

Effect Size	Sample Size
-0.20	225
-0.17	365
-0.10	945
-0.02	26100
0.05	4600
0.10	1250