Power Calculation for Off-grid and Grid

Xiaoming Zhang

2023-10-04

Formula used for power calculation

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \boxed{\left[1 + \rho(m-1)\right]}$$

Figure 1: Clustered sampling power calculation of sample size

Wellbeing

Explanation

Standard deviation and minimum detectable effect come from the Lee et al. paper. Specifically the wellbeing measurement is in panel D: primary non-economic outpus- **D2 normalized life satisfaction**.

- Standard deviation: Because the **D2 normalized life satisfaction** question was not gathered in baseline, standard deviation of the population was estimated as 1 in the paper.
- Minimum detectable effect: The mdf was obtained using regression results for treatment effects on key outcomes (D2 in this case). As it is shown as the table below, therefore we're trying numbers ranging from 10% to 20% for this exercise.

Table 1: Well being regression results from Lee et. al

Category	ITT	TOT
round1	0.11	0.12
round2	0.22	0.29
pooled	0.16	0.19

Power calculation

The number of villages surveyed (number of clusters) are set at 50.

This is the code used to calculate the number of households needed to be surveyed in order to see certain percent increase in reported life satisfaction(on a scale of 1-10)

```
###10%

rho <- 0.05
z <- 2.80 #alpha =  0.05, beta = 0.8

signma <- 1 #paper assumed normal distribution because the wellbeing question was not asked baseline
signma_sqr <- signma^2

mdf1 <- 0.10
mdf1_sqr <- mdf1^2

m <- 50 #number using villages in newest EPC scope

n1 <- ((4 * z * signma_sqr) / mdf1_sqr) * (1 + rho * (m - 1))
```

Results

Table 2: Well being power calculations

MDF	Households
10%	3864
16%	1509
19%	1070

Energy Spending

Explanation

Standard deviation and minimum detectable effect come from the Lee et al. paper. Specifically the energy spending measurement is in panel B: additional energy outcomes- **B8. Monthly total energy spending(USD)** & **B7. Monthly kerosene spending(USD)**.

B8. Monthly total energy spending(USD)

- Minimum detectable effect: Due to the fact that the coefficient for **B8. Monthly total energy spending(USD)** is not statistically significant, and the spending increase is around 3%-5%, the mdf used in the power calculation is calculated at 5%, 10%, 20% level of decrease from the control group.

```
library(knitr)

data1 <- data.frame(
  Category = c("round1", "round2", "pooled"),
  Control = c(11.66, 9.98, 10.83),
  ITT = c(-0.35, -0.48, -0.36),
  TOT = c(-0.40, -0.01, -0.19)
)

kable(data1, format = "pipe", caption = "Monthly total energy spending(USD) regression results from Lee
```

Table 3: Monthly total energy spending(USD) regression results from Lee et. al

Category	Control	ITT	TOT
round1	11.66	-0.35	-0.40
round2	9.98	-0.48	-0.01
pooled	10.83	-0.36	-0.19

Power calculation

The number of villages surveyed (number of clusters)are set at 50.

This is the code used to calculate the number of households needed to be surveyed in order to see certain USD decrease in monthly total energy spending

```
###5%

rho <- 0.05
z <- 2.80 #alpha =  0.05, beta = 0.8

signma_spending1 <- 21.83 # Got directly from the Lee et al. paper
signma_spending1_sqr <- signma_spending1^2

mdf_spending1 <- 10.83*0.05 #pulled control*5% decrease
mdf_spending1_sqr <-mdf_spending1^2

m <- 50 #number using villages in newest EPC scope

n1_spending <- ((4 * z * signma_spending1_sqr) / mdf_spending1_sqr) * (1 + rho * (m - 1))
```

Results

Table 4: Montly total energy spending(USD) power calculations

MDF	Households
-5%	62798
-10%	15700
-20%	3925

B7. Monthly kerosene spending(USD)

Because the decrease of spending on total energy is not statistically significant, hereby runs the power calculation for monthly kerosene spending, which is statistically significant at 99%

- mdf: From the regression, it seems that the decrease in kerosene spending range from 30% TO 40%, therefore for the mdf in this analysis, I'm using 30%, 35%, and 40% decrease in kerosene spending

Table 5: Monthly total kerosene spending(USD) regression results from Lee et. al

Category	Control	ITT	TOT
round1	2.81	-1.15	-1.20
round2	2.47	-0.63	-0.74
pooled	2.64	-0.90	-1.00

Power calculation

The number of villages surveyed (number of clusters)are set at 50.

This is the code used to calculate the number of households needed to be surveyed in order to see certain USD decrease in monthly kerosene spending

```
###30%

rho <- 0.05
z <- 2.80 #alpha =  0.05, beta = 0.8

signma_kerosene1 <- 2.75 # Got directly from the Lee et al. paper
signma_kerosene1_sqr <- signma_kerosene1^2

mdf_kerosene1 <- 2.64*0.3 #pulled control*5% decrease
mdf_kerosene1_sqr <-mdf_kerosene1^2

m <- 50 #number using villages in newest EPC scope

n1_kerosene <- ((4 * z * signma_kerosene1_sqr) / mdf_kerosene1_sqr) * (1 + rho * (m - 1))
```

Results

Table 6: Montly total kerosene spending(USD) power calculations

MDF	Households
-30%	466
-35%	342
-40%	262

Electricity Consumption, very primary thought.

Explanation

Standard deviation and minimum detectable effect come from the Lee et al. paper. Specifically the electricity consumption measurement is in **Table B7: Benchmarking average monthly electricity consumption in kWh and USD - Panel A: Newly connected households in sample**.

- Standard deviation: we are given Q1 and Q3, the standard deviation is calculated using the formula below. Sepcifically the standard deviation is taking the IQR of households that are newly connected in the sample and their consumption in 2016 (Round 1 survey)

$$\text{IQR} = Q3 - Q1$$

$$\text{Standard Deviation} \approx \frac{\text{IQR}}{1.349}$$

Figure 2: IQR estimation of sd

- Minimum detectable effect is estimated at 30%, 35%, 40% from the calculated increase rate in electricity consumption from 2016 to 2017 in newly connected households.
- The data used could be found on ***page 101*** of the Lee etal. pdf (A-57 in terms of page number)

Power calculation

The number of villages surveyed (number of clusters)are set at 50.

This is the code used to calculate the number of households needed to be surveyed in order to see certain increase in electricity consumption

```
###30%

rho <- 0.05
z <- 2.80 #alpha =  0.05, beta = 0.8

signma_consumption1 <- s2 #s2 is the calculated standard deviation
signma_consumption1_sqr <- signma_consumption1^2

mdf_consumption1 <- 7.9*0.3 #treatment*increase??? Not sure here
mdf_consumption1_sqr <-mdf_consumption1^2

m <- 50 #number using villages in newest EPC scope

n1_consumption <- ((4 * z * signma_consumption1_sqr) / mdf_consumption1_sqr) * (1 + rho * (m - 1))
```

Results

Table 7: Average montly electricity consumption in kWh

MDF	Households
30%	1830
35%	1344
40%	1029