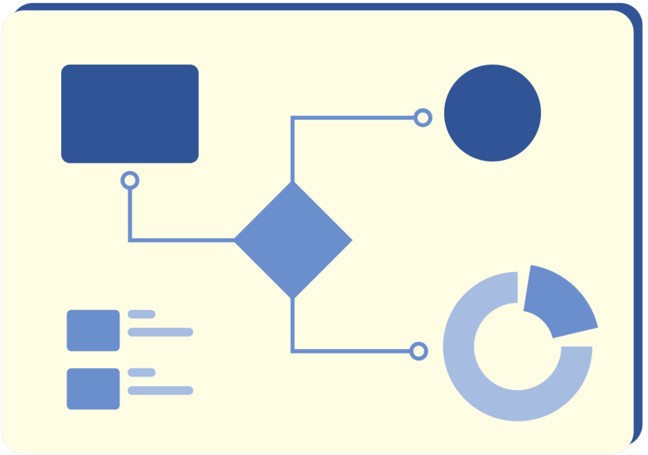
Random sample size for single and multiple hypothesis tests

Z-test proportions: two-samples one-tailed



June 2024

Overview

1. Single hypothesis test
2. Multiple hypothesis tests
3. References
4. Code

Activities NGO Algorithm Audit

Normative advice commissions — Advising on ethical issues that arise in concrete algorithmic practice through deliberative and diverse normative advice commissions, resulting in algoprudence

Technical tools — Implementing and testing technical tools to detect and mitigate bias, e.g., bias detection tool and synthetic data generation

Knowledge platform — Bringing together experts and knowledge to foster the collective learning process on the responsible use of algorithms, e.g., AI Policy Observatory and white papers

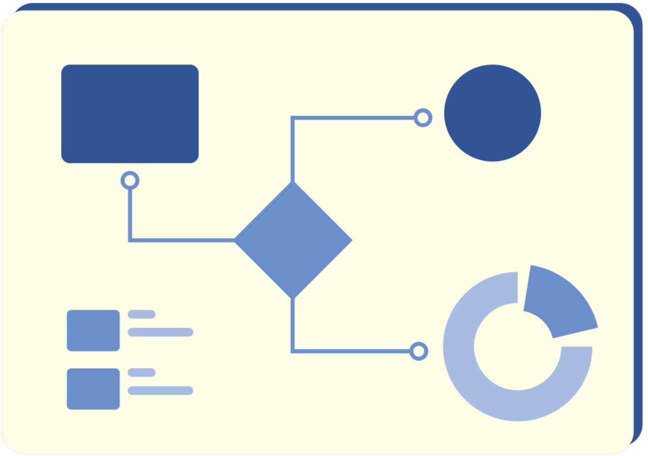
Supported by

SIDNfonds

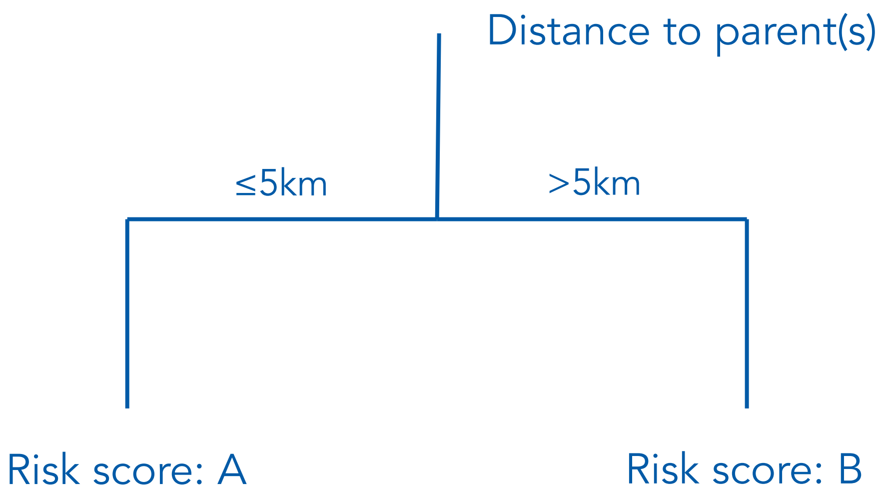
European Artificial Intelligence & Society Fund

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Algorithm Audit



Risk profile consitsting of single criterion



Theoretical answer

- **Goal:** Determine minimum random sample size to measure statistically significant difference in unduly use rate between group A and group B
- **Data needed:**
  - p_A: proportion unduly group A, i.e. $p_A = \frac{N_{unduly,A}}{N_A}$, where N_A is the sample size of group A
  - p_B: proportion unduly group B, i.e. $p_B = \frac{N_{unduly,B}}{N_B}$, where N_B is the sample size of group B
  - Null and alternative hypothesis (one-sided):

    H_0: $p_A = p_B$
    H_1: $p_A > p_B$
  - Significance level (**α**): probability of false positive (accepting H_0 while H_1 is true), e.g., 0.05 or 0.01
  - Power **(1 - β)**: probability of true positive (rejecting H_0 while H_0 is indeed false), e.g., 0.8 or 0.9
  - Ratio of sample size (k): ratio of sample size group A and sample size group B, i.e. $\mathrm{k} = \frac{N_A}{N_B}$

Theoretical answer

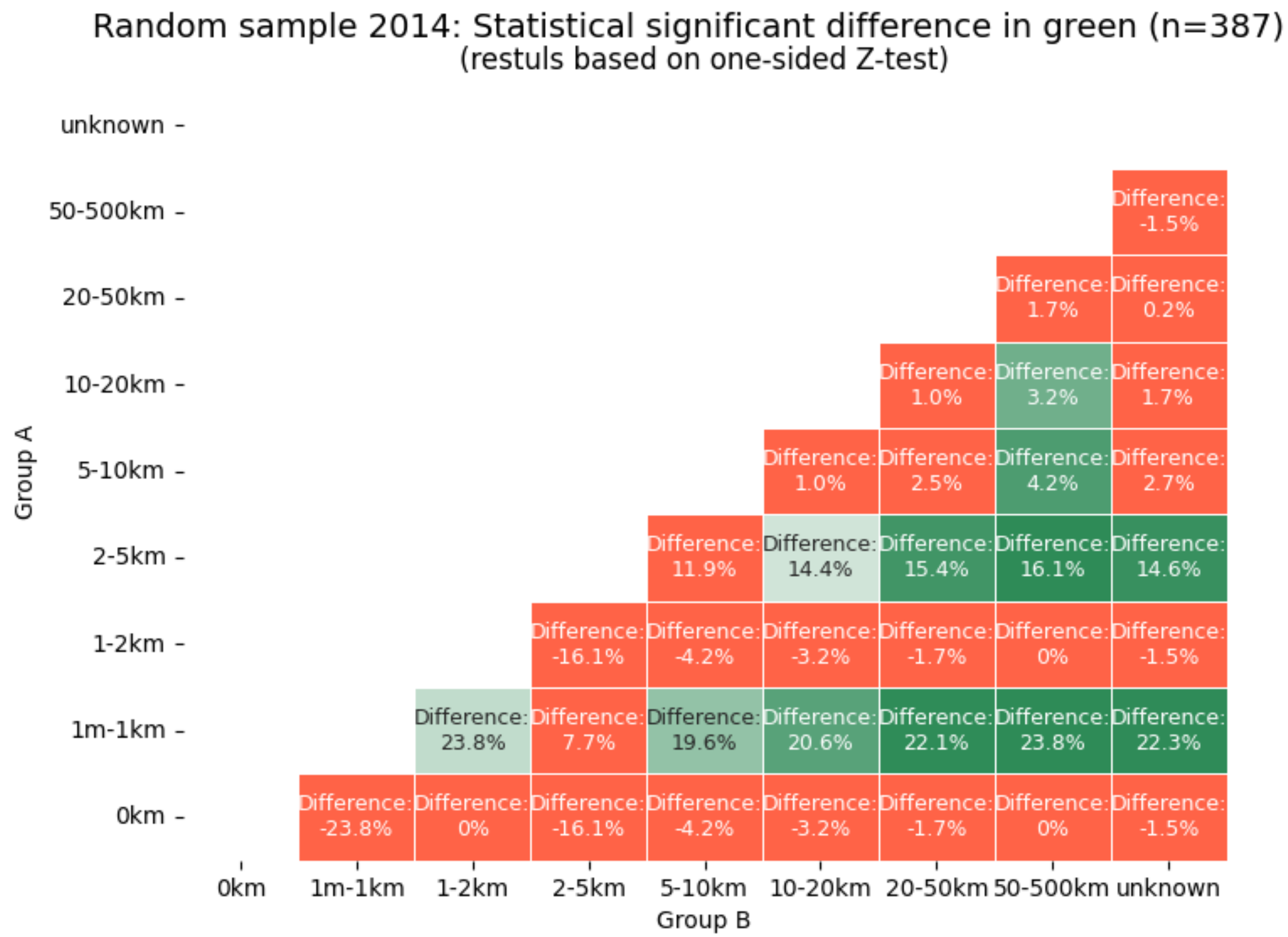
> Formula to determine random sample size (n):

$$n = \left(\frac{p_A(1-p_A)}{k} + p_B(1-p_B)\right) \cdot \left(\frac{(z_{1-\alpha}+z_{1-\beta})}{(p_A-p_B)}\right)^2, \text{ where}$$

$z_{1-\alpha}$:critical value of the normal distribution at the $1-\alpha$ confidence level

$z_{1-\beta}$:critical value of the normal distribution at the $1-\beta$ confidence level

Example: Distance data DUO random sample 2014 (n=387)



Example: Distance data DUO random sample 2014 (n=387)

Sample 2014	Size of group	# unduly grants	Percentage
0km	8	0	0%
1m-1km	21	5	23,8%
1-2km	11	0	0%
2-5km	31	5	16,1%
5-10km	24	1	4,2%
10-20km	31	1	3,2%
20-50km	58	1	1,7%
50-500km	137	0	0%
onbekend	66	1	1,5%
Totaal	387	14	3,6%

Group A
$p_A = 14\%$
$N_A = 71$

Group B
$p_B = 1{,}3\%$
$N_B = 316$

Table 9 – Overview of group size and unduly use percentage per distance category in the 2014 random sample (n=387).

Example: selection criterion 'distance to parent(s)' group A: ≤5km, group B: >5km

> $p_A = 0.013$ (guestimation)

> $p_B = 0.14$ (guestimation)

> Hypothesis test:

H_0: $p_A=p_B$
H_1: $p_A>p_B$

> Significance level (α): 0.05

> Power (1 - β): 0.8

> Ratio of sample size (k): $\frac{71}{316} = 0.225$ (based on random sample), $\frac{45.79k}{202.87k} = 0.226$ (based on full population sample)

> Random sample size for group B: 69

> Random sample size for group A: 0.225*69=16

--> Random sample size: 85

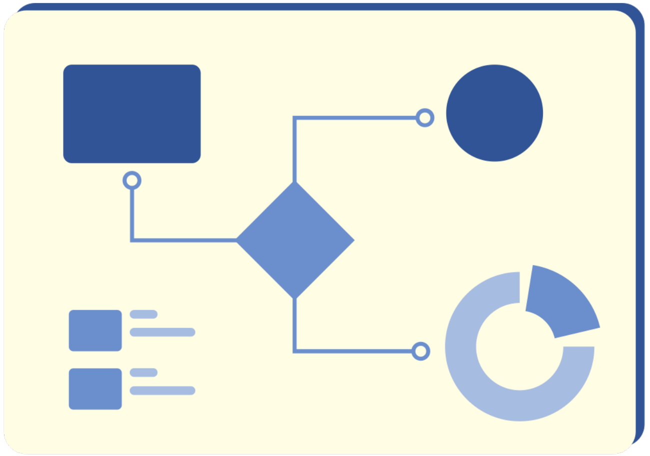
Algorithm Audit

Varying confidence level, power and percentual difference influences sample size

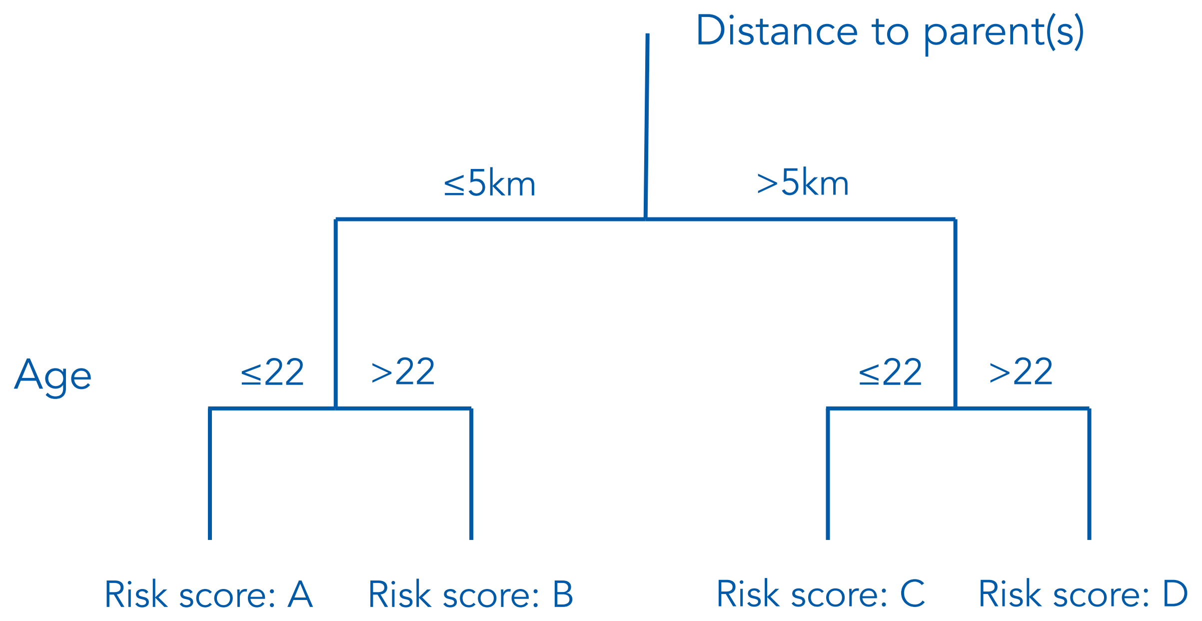
		p_A, p_B			
α	(1 - β)	1%, 2%	1%, 7%	1.3%, 14%	2%, 20%
0.05	0.8	N_A = 3.933, N_B = 885 n=4.818	N_A = 188, N_B = 43 n=231	N_A =16, N_B =69 n=85	N_A = 48, N_B = 11 n=59
	0.9	N_A = 5.477, N_B = 1.226 n=6.673	N_A = 260, N_B = 59 n=319	N_A = 95, N_B = 22 n=117	N_A = 66, N_B = 15 n=81
0.01	0.8	N_A = 6.383, N_B = 1.437 n=7.820	N_A = 305, N_B = 69 n=374	N_A = 111, N_B = 25 n=136	N_A = 77, N_B = 18 n=95
	0.9	N_A = 8.279, N_B = 1.863 n=10.142	N_A = 395, N_B = 89 n=484	N_A = 144, N_B = 33 n=177	N_A = 100, N_B = 23 n=123

Note:
- Higher confidence --> larger sample size
- Measuring smaller difference --> larger sample size

Algorithm Audit



Risk profile consitsting of two criteria



Theoretical answer to compute random sample size for m=2 hypothesis tests

- **Goal:** Determine minimum random sample size to measure statistically significant difference in unduly use rate between students living ≤5km (group A) and >5km away (group B), and students ≤22 years old (group C) and >22 years old (group D)
- **Data needed:**
  - p_A, p_B: proportion unduly group ≤5km and proportion unduly group >5km
  - p_C, p_D: proportion unduly group ≤22 years old and proportion unduly group >22 years old

  - Two null and alternative hypothesis (one-sided):

    H_0: $p_A = p_B$ H_0: $p_C = p_D$
    H_1: $p_A > p_B$ H_1: $p_C > p_D$
  - Significance level (**α**): probability of false positive (accepting H_0 while H_1 is true), e.g., 0.05 or 0.01
  - Bonferroni correction for testing m=2 hypothesis: $\alpha_{adjusted} = \frac{\alpha}{2}$
  - Power **(1 - β)**: probability of true positive (rejecting H_0 while H_0 is indeed false), e.g., 0.8 or 0.9
  - Ratio of sample sizes (k_1 and k_2): ratio of group sample sizes, i.e. $k_1 = \frac{N_A}{N_B}, k_2 = \frac{N_C}{N_D}$

Formula to compute sample size for m=2 hypothesis tests

> Formula to determine random sample size (n_1) to test first hypothesis test and random sample size (n_2) to test second hypothesis test:

$$n_1 = \left(\frac{p_A(1-p_A)}{k_1} + p_B(1-p_B)\right) \cdot \left(\frac{\left(z_{1-\alpha_{adjusted}} + z_{1-\beta}\right)}{(p_A - p_B)}\right)^2,$$

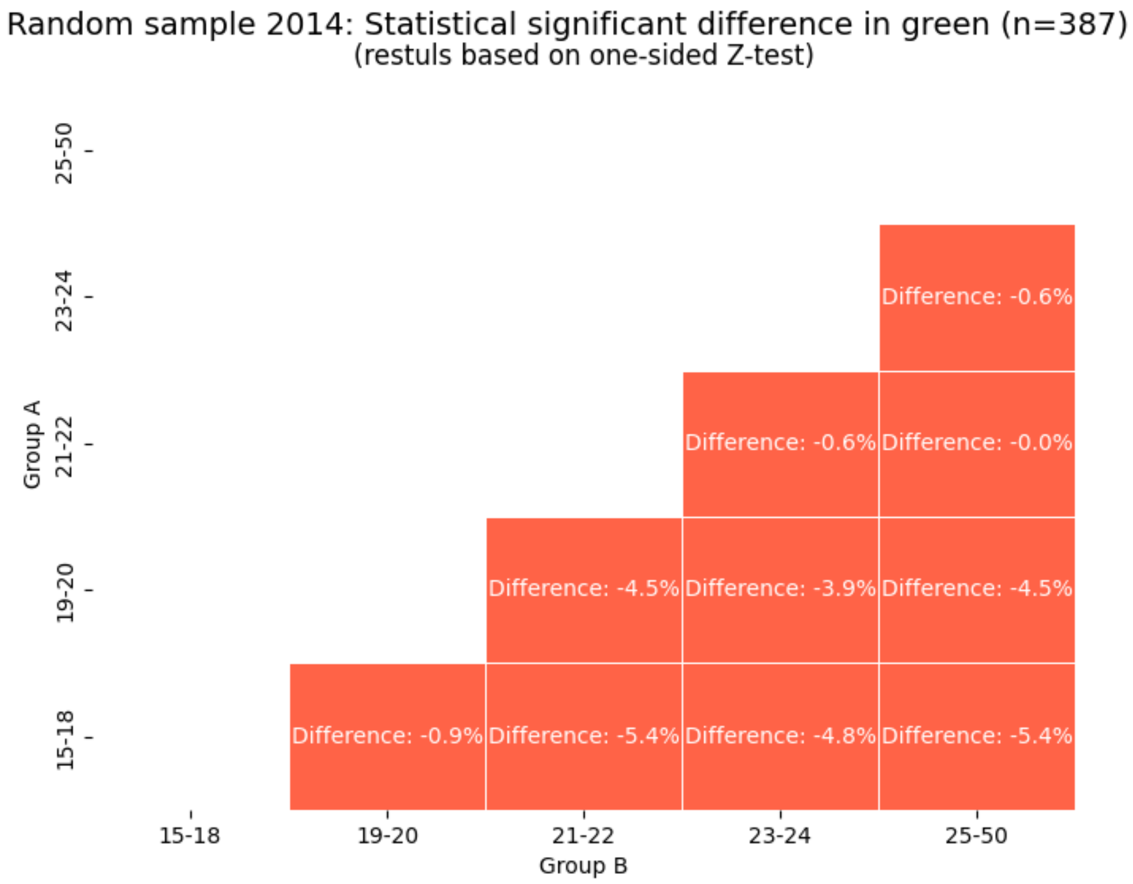
$$n_2 = \left(\frac{p_C(1-p_C)}{k_2} + p_D(1-p_D)\right) \cdot \left(\frac{\left(z_{1-\alpha_{adjusted}} + z_{1-\beta}\right)}{(p_C - p_D)}\right)^2,$$

where:

$z_{1-\alpha_{adjusted}}$: critical value of the normal distribution at the $\alpha_{adjusted} = \frac{\alpha}{2}$ confidence level

$z_{1-\beta}$: critical value of the normal distribution at the $1-\beta$ confidence level

Example: Age data DUO random sample 2014 (n=387)



Example: Age data DUO random sample 2014 (n=387)

Sample 2014	Size of group	# unduly grants	Percentage
15-18	24	0	0%
19-20	115	1	0,9%
21-22	149	8	5,4%
23-24	62	3	4,8%
25-50	37	2	5,4%
Totaal	387	14	3,6%

Table 7 – Overview of group size and unduly use percentage per age category in the random sample 2014 (n=387).

Group C
$p_C = 3{,}2\%$
$N_C = 288$

Group D
$p_D = 5{,}0\%$
$N_D = 99$

Algorithm Audit

Example: testing m=2 hypotheses for distance ≤5km vs 5km and age ≤22 vs >22 years

> $p_A = 0.013, p_B = 0.14$ (guestimation)

> $p_C = 0.032, p_D = 0.05$ (guestimation)

> Hypothesis test:

H_0: $p_A = p_B$ H_0: $p_C = p_D$
H_1: $p_A > p_B$ H_1: $p_C > p_D$

> Significance level (α): 0.05/2=0.025

> Power (1 - β): 0.8

> Ratio of sample sizes: $k_1 = \frac{71}{316} = 0.225$ and $k_2 = \frac{288}{387} = 0.744$ (based on random sample)

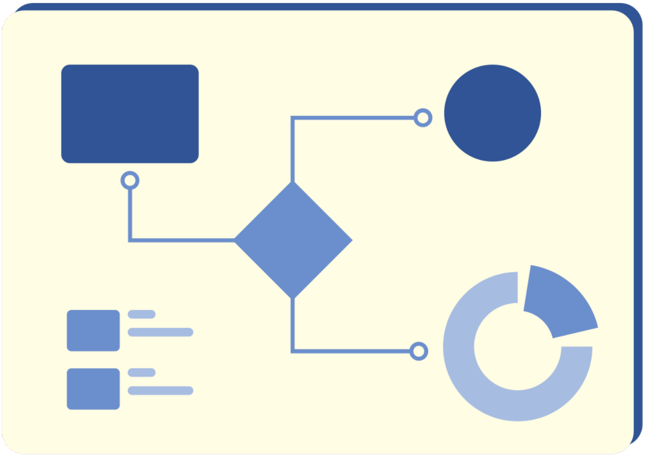
> Random sample size for group B: 69
> Random sample size for group A: 0.225*69=16

> Random sample size for group D: 2.160
> Random sample size for group C: 0.744*2.160=1.607

Generalizable to testing m hypotheses

--> Random sample size: 85+3.767=3.872

Algorithm Audit



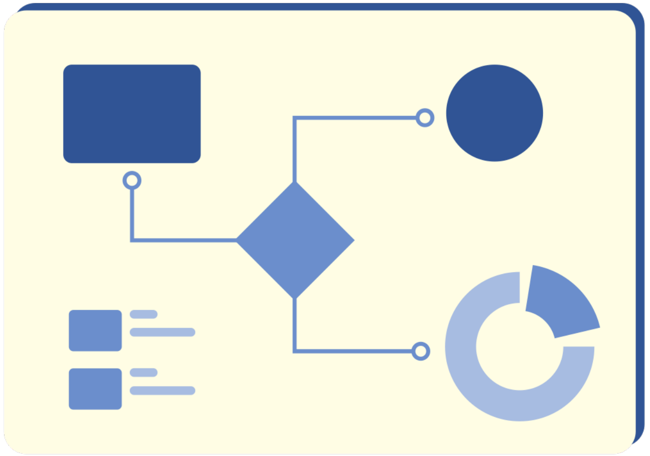
References

> http://powerandsamplesize.com/Calculators/Compare-2-Proportions/2-Sample-1-Sided

> https://stats.stackexchange.com/questions/581679/minimum-sample-size-calculations-for-two-sample-two-tailed-proportion-z-test

> https://stats.stackexchange.com/questions/467595/how-do-you-calculate-sample-sizes-for-multiple-treatments?noredirect=1&lq=1

Algorithm Audit



Python code

```
# Define the parameters
def sample_size(p, k, alpha, beta, m):
    pA = p[0]
    pB = p[1]
    kappa = k
    alpha = alpha
    beta = beta
    print("pA:", pA)
    print("pB:", pB)
    print("alpha:", alpha)
    print("beta:", beta)

    # Calculate the z-scores for the given alpha and beta
    z_alpha = stats.norm.ppf(1 - alpha/m)
    z_beta = stats.norm.ppf(1 - beta)

    # Calculate the sample size nB using the given formula
    nB = (pA * (1 - pA) / kappa + pB * (1 - pB)) * ((z_alpha + z_beta) / (pA - pB))**2

    # Round up nB to the nearest whole number
    nB_ceiling = np.ceil(nB)
    nA_ceiling = np.ceil(kappa * nB)

    n = nA_ceiling + nB_ceiling

    # Print the ceiling value of nB
    print("Sample size in group B:", int(nB_ceiling))
    print("Sample size in group A:", int(nA_ceiling))
    print("Sample size needed:", int(n))

    return(n)
```

Algorithm Audit

www.algorithmaudit.eu

info@algorithmaudit.eu

https://www.linkedin.com/company/algorithm-audit/

https://github.com/NGO-Algorithm-Audit

Stichting Algorithm Audit is bij de Kamer van Koophandel
geregistreerd onder nummer 83979212