

Project 2 – your title here

MTH 161 – Fall 2024

Emma Becker

2024-11-15

Part 1: choose a dataset and propose a question

Due Nov. 15th

Link to data source: <https://www.internationalgenome.org/data-portal/sample>

The dataset I chose was this one found under the “Awesome Public Datasets” link and more specifically under the Biology tab of this source provided in the assignment resources. The dataset is titled “ISGR: The International Genome Sample Resource: Supporting Open Human Variation Data”. The dataset meets the conditions necessary, it has over 100 observations (more specifically it has 4,978 observations and displays 10 at a time). There are also at least 5 column variables which include 15 different types of genes that the participants may have. Additional description variables include the sex, population ethnicity, and a label for each participant.

The categorization of either having or not having 1000 genomes on each gene variable is being measured to investigate any possible links between disease on a certain gene and the population of the individual.

My research question that I will investigate using this dataset is:

“Is there a significant link between having 1000 genomes on the GRCh38 gene and being both female and Puerto Rican in Puerto Rico?”

The parameter of interest (p) is the entire population of Puerto Rican female individuals living in Puerto Rico who have 1000 genomes on the GRCh38 gene.

The population statistic (\hat{p}) is the number of observed female individuals living in Puerto Rico who do have 1000 genomes on this gene.

The null hypothesis is that there is no significant difference between this population having 1000 genomes on this specific GRCh38 gene as compared to other populations, $P = 0$.

The alternative hypothesis is that there is a significant difference between this population have 1000 genomes on this specific GRCh38 gene as compare the other populations. P does not equal 0.

i Note

The instructions above are included here for your convenience, but they should not appear in your final report.