**Improving privacy guarantee on health data using RAPPOR**

**Meenakshi** **Nagarajan**

**Abstract**

Data obtained from users through surveys or questionnaires are major source of information for machine learning & analytics operations. Results of analysis performed on the data can help obtain new insights on product design. Application of different statistical queries on the obtained data should not reveal the identity of the individual to avoid privacy attacks. However, publishing such statistical data on public platform may lead to inference attack on database. Any updates to the published data, may result in difference in the statistical parameters between old and new data. This may cause background knowledge attack. Often with manipulative queries and SQL injection, attacker can identify a single user from statistical database which may lead to inferential attacks. The proposed approach of using RAPPOR mechanism can preserve user’s identity by adding random noise to the actual data.

1. **INTRODUCTION**

Big Data, Machine learning and cloud computing in state-of-the-art world is being considered as the new opportunity for analytics (Begum & Nausheen, 2018). With rise in the number of online communities for data scientists and machine learners, several datasets are published online to provide a platform for engineers to learn to analyze data and build best models for solving data specific challenges. Publishing such statistical data on public platform may lead to privacy attacks on data and result in exposing sensitive information specific to any individual. (Dwork, McSherry, Nissim, & Smith, 2006). Statistical analysis on user data is a potential method to obtain new insights on any product or process design which eventually improves the user experience. Therefore, performing statistical analysis on user data or user feedback about a product and maintaining user’s privacy are equally important.

The database is assumed to be present on a trusted server with authorization and access controls on sensitive data. Statistical data which is to be published for common use will be taken from this trusted database by writing a query which will aggregate user statistics. To protect privacy of the user, a noise can be added at the end of the obtained results. But it can be easily identified by attackers if the noise is a fixed value. Adding Laplacian noise or gaussian distributed noise over the dataset are some well-known techniques which are used only in theory may improve privacy at the cost of accuracy. The goal of this project is to use RAPPOR differential privacy technique on mental illness data to preserve privacy and improve accuracy of the published information.

1. **RELATED WORK**

Often data collection from users preserves privacy by not collecting any sensitive information about the individual. However, using different statistical queries and aggregating the results obtained can reveal some knowledge about the user compromising the privacy. Using epsilon differential privacy technique by adding Laplacian noise on the input data will result in increased privacy (Dwork, McSherry, Nissim, & Smith, 2006). On a typical survey, the data may be categorical. Adding Laplacian noise on a categorical data will result in loss of semantic information. Adding exponential noise on client’s side might not help the analysts to do data sampling (McSherry & Talwar, 2007). Using dimensionality reduction techniques like l-anonymity and k-diversity may result in increased privacy but utility of the information published is compromised (Liu, Jiang, Sha, & Govindan, 2012).

Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR) is a technique which improves accuracy by preserving crowdsourced information obtained from surveys, questionnaires, feedback etc. Also, on randomizing the response from participants and approximating the probability of correctness in response, privacy of the user is preserved. Applying this technique on any data would significantly contribute to accuracy and privacy of published information (Erlingsson, Pihur, & Korolova, 2014).

1. **MY APPROACH**

Mental disease data surveys were conducted across different communities working in several industrial sectors (https://www.kaggle.com/osmi/mental-health-in-tech-survey). This data is open to raising awareness, educating and providing resources to support mental well - being in technology and open source communities. Using these data, researchers intend to find the effect of the frequency of mental health diseases and attitudes to mental health varying by geographical location. Data from one geographical region which involves a small range of audience knowing each other, there are many changes for the attacker to find details about an individual. Publishing this mental health data on a public platform can result in privacy concerns. We will be using differential privacy technique, RAPPOR, on a real-world data collection to demonstrate how it preserves privacy and maintain statistical accuracy of the data.

RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response) (Erlingsson, Pihur, & Korolova, 2014) uses Bloom filter based randomized response technique on strings. Once the data has been collected in the bloom filter, two levels of noise will be introduced on the data. The first step is Permanent randomized response that is used to create a noisy answer. It also uses memoization technique to memoize the noisy answer unique to an individual and to reuse it in place of the real answer. The second step is Instantaneous randomized response, to report the result of noisy answer. These results are then sent to the server. The large amount of randomness in the response makes sure that the operator cannot draw meaningful conclusions about the user. The concept of RAPPOR was introduced by Google to improve learn the statistics of user data by preserving user privacy. In this project, we use RAPPOR differential privacy technique on mental illness data to find if it preserves privacy of user data and improve accuracy of the statistics of the data.

1. **EXPERIMENTAL EVALUATION: EVALUATION GOALS, DESIGN, DATASETS, SETUP, AND RESULT ANALYSIS**

**4.1 Evaluation goals**

To design scenarios to demonstrate the effect of different privacy attacks on input data and evaluate the use of RAPPOR technique on mental health tech survey data in preserving the privacy and maintain the statistical accuracy of the data.

**4.2 Design and Dataset**

Dataset for this project was obtained from <https://www.kaggle.com/osmi/mental-health-in-tech-survey>. R queries will be the primary mode to access data from the database. To identify individuals on data I employed inferential attack and background knowledge attack. I designed specific scenarios to better understand the consequence of the attacks.

# Obtaining Input:

mental\_health\_data<-read.csv("/Users/meenakshinagarajan/Desktop/Privacy aware computing/survey.csv", header=TRUE, sep=",")  
head(mental\_health\_data)

## Timestamp Age Gender Country state self\_employed  
## 1 2014-08-27 11:29:31 37 Female United States IL <NA>  
## 2 2014-08-27 11:29:37 44 M United States IN <NA>  
## 3 2014-08-27 11:29:44 32 Male Canada <NA> <NA>  
## 4 2014-08-27 11:29:46 31 Male United Kingdom <NA> <NA>  
## 5 2014-08-27 11:30:22 31 Male United States TX <NA>  
## 6 2014-08-27 11:31:22 33 Male United States TN <NA>  
## family\_history treatment work\_interfere no\_employees remote\_work  
## 1 No Yes Often 6-25 No  
## 2 No No Rarely More than 1000 No  
## 3 No No Rarely 6-25 No  
## 4 Yes Yes Often 26-100 No  
## 5 No No Never 100-500 Yes  
## 6 Yes No Sometimes 6-25 No  
## tech\_company benefits care\_options wellness\_program seek\_help  
## 1 Yes Yes Not sure No Yes  
## 2 No Don't know No Don't know Don't know  
## 3 Yes No No No No  
## 4 Yes No Yes No No  
## 5 Yes Yes No Don't know Don't know  
## 6 Yes Yes Not sure No Don't know  
## anonymity leave mental\_health\_consequence  
## 1 Yes Somewhat easy No  
## 2 Don't know Don't know Maybe  
## 3 Don't know Somewhat difficult No  
## 4 No Somewhat difficult Yes  
## 5 Don't know Don't know No  
## 6 Don't know Don't know No  
## phys\_health\_consequence coworkers supervisor mental\_health\_interview  
## 1 No Some of them Yes No  
## 2 No No No No  
## 3 No Yes Yes Yes  
## 4 Yes Some of them No Maybe  
## 5 No Some of them Yes Yes  
## 6 No Yes Yes No  
## phys\_health\_interview mental\_vs\_physical obs\_consequence comments  
## 1 Maybe Yes No <NA>  
## 2 No Don't know No <NA>  
## 3 Yes No No <NA>  
## 4 Maybe No Yes <NA>  
## 5 Yes Don't know No <NA>  
## 6 Maybe Don't know No <NA>

From this large dataset, we are choosing only a random sample of rows and columns that are vulnerable to privacy attacks for the purpose of demonstration. We consider columns 'Age', 'Gender', 'Country', and 'treatment' from this dataset as potential elements that are vulnerable to attacks. Below are the two scenarios which describes the possibility of privacy attacks, when the data is exposed to public. The attacker could gain knowledge on user and his/her mental health treatment in below scenarios even though their anonymity is protected.

# Inference attack: Finding if an individual has mental health issue

Scenario 1: Revealing Time Critical Survey Data

A survey company is planning to take a survey regarding user's mental health across regions to construct mental health awareness camps. While taking survey, they have decided to publish health related data excluding user's private information. They have decided to tour Company A for 2 days. Let the person who is answering the survey have health issues and is the attacker.

#Day 1 survey  
myvars <- c("Age", "Gender", "Country","treatment")  
newdata <- mental\_health\_data[myvars]  
mysample <- newdata[11:20,]  
head(mysample)

## Age Gender Country treatment  
## 11 31 Male United States Yes  
## 12 29 male Bulgaria No  
## 13 42 female United States Yes  
## 14 36 Male United States No  
## 15 27 Male Canada No  
## 16 29 female United States Yes

#Selecting observations from day 1 survey where mental treatment is 'Yes'  
mysample\_yes <- mysample[ which(mysample$treatment=='Yes'),]  
  
#number of female who answered 'Yes'  
print("Number of females who took the mental treatment in day 1 survey:")

## [1] "Number of females who took the mental treatment in day 1 survey:"

nrow(mysample\_yes[ which(mysample\_yes$Gender=='female'), ])

## [1] 2

#Day 2 survey  
mynewsample <- newdata[31:40,]  
  
#Selecting observations from day 2 survey where mental treatment is 'Yes'  
mynewsample\_yes <- mynewsample[ which(mynewsample$treatment=='Yes'),]

#Selecting observations from day 2 survey where mental treatment is 'Yes' an Gender is 'Female'  
print("Number of females who took the mental treatment in day 2 survey:")

## [1] "Number of females who took the mental treatment in day 2 survey:"

nrow(mynewsample\_yes[ which(mynewsample\_yes$Gender=='female'), ])

## [1] 1

The total number of female respondents who took the mental health treatment on day 1 and day 2 together is 3. If the attacker is one among the survey respondents and he has knowledge on statistics of day 1 survey, then on combining the data obtained from day 1 and day 2, he can easily conclude that only one female respondent attended day 2 survey and he could access her mental treatment data.

# Background Knowledge attack: Finding if an individual has mental health issue

Scenario 2: Company has only 1 employee is equal to or above 50

mynewsample

## Age Gender Country treatment  
## 31 32 Male United Kingdom No  
## 32 31 Male United States No  
## 33 30 male United Kingdom Yes  
## 34 42 Male United States Yes  
## 35 40 female United States Yes  
## 36 27 Male United States Yes  
## 37 29 Male Canada No  
## 38 38 Male Portugal No  
## 39 50 M United States No  
## 40 35 M United States Yes

print("Number of persons who said 'Yes' to treatment in day 2 survey and 50 years of age:")

## [1] "Number of persons who said 'Yes' to treatment in day 2 survey and 50 years of age:"

nrow(mynewsample[ which(mynewsample$Age=='50'), ])

## [1] 1

On Publishing this data people inside company, who has knowledge about their peers, can find out the private field of the individual.

**4.3 Setup**

This source code has been adopted from <https://github.com/google/rappor> . For Simulation, there are specific files and functions which are used.

In the folder R/analysis, there are functions for encoding and decoding a particular client values. Change the function SetOfStrings in R/analysis/simulation.R file

SetOfStrings <- function(num\_strings = 100) {

# Generates a set of strings for simulation purposes. #strs <- paste0("V\_", as.character(1:num\_strings)) strs <-

c("30to40MaleUSNo","30to40MaleUSYes","40to50MaleUSYes","40to50MaleUSNo","20to30MaleUSNo","20to30MaleUSYes","30to40MaleUKNo","30to40MaleUKYes","40to50MaleUKYes","40 to50MaleUKNo","20to30MaleUKNo","20to30MaleUKYes", "30to40MaleCanadaNo","30to40MaleCanadaYes","40to50MaleCanadaYes","40to50CanadaUK No","20to30MaleCanadaNo","20to30MaleCanadaYes")

strs

}

**Experimental evaluation parameters**

In console load these parameters,

params\_4x2 <- list(k = 16, m = 8, h = 2,p=0.5,q=0.75,f=0.5)

k = Number of bits

m = Number of cohorts

h = Number of Hash Functions

p,q = Probability values

f = Frequency

**Inputs**

Sample Size: 10000 user survey responses are generated.

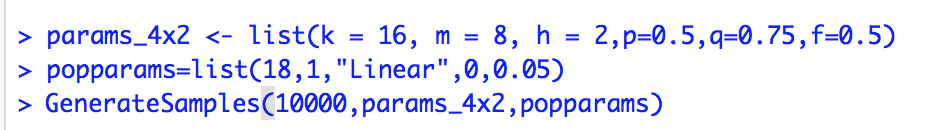


Fig. 1 Console inputs

**Implementation and Outputs**

The implementation is done to demonstrate effectiveness of RAPPOR technique in preserving the privacy of individual data from the survey. The first step towards to implementation is to convert strings to bit values using hashing techniques. These bit values represent data from an individual. Bloom filters are applied to the user inputs which is followed RAPPOR encoding. This encoded data and parameters such as h, k m, p, q, f, used for encoding are sent to server.

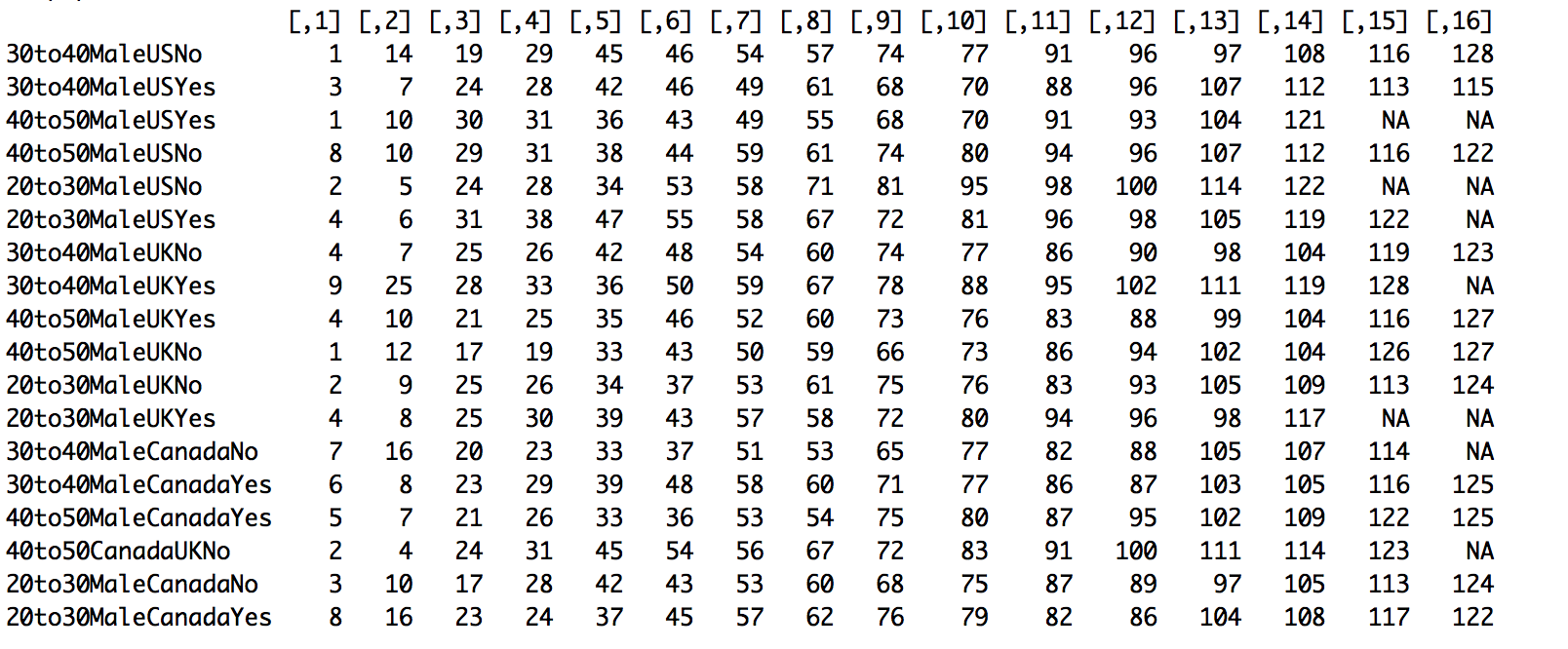


Fig. 2 Bloom filter bits vs. input – Bloom filter is represented by 16 bits (k). It means each row of the survey is represented as 16 bits.

GenerateSamples method mentioned in the screenshot (Fig. 1) generates 10000 samples for user values from survey and perform encoding on the data. Results of RAPPOR encoding is given in below screenshot (Fig.3). It is a matrix of 17 candidate strings and 8 cohorts. Each value in the matrix represents rappor encoded values for each candidate string.

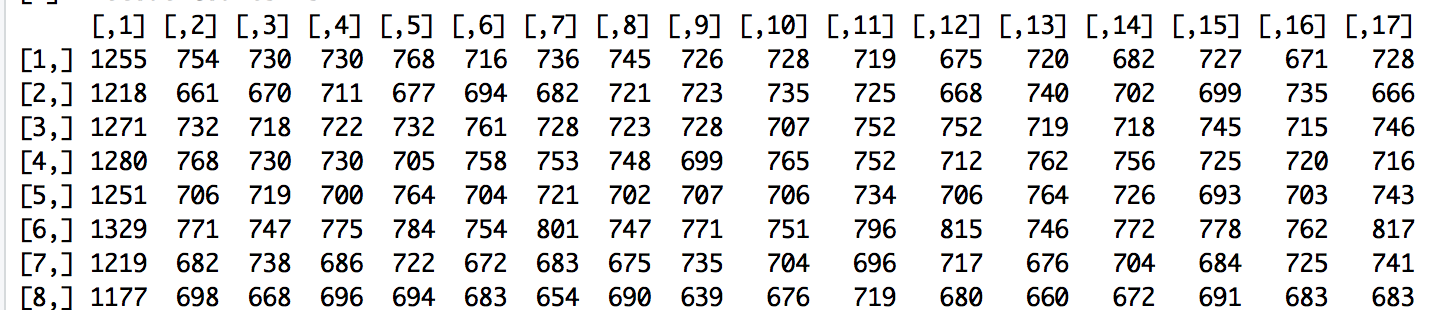


Fig. 3 RAPPOR Encoded data

This technique can preserve data privacy by sending the encoded string to server and not the actual data. Now that all the values are encoded, it is followed by decode function which could be used for publishing and for further statistical analysis.

RAPPOR encoded values, Bloom filter map (Fig. 2), parameters (Fig. 1) are sent to the decode function. The obtained results are given in the below screen shots

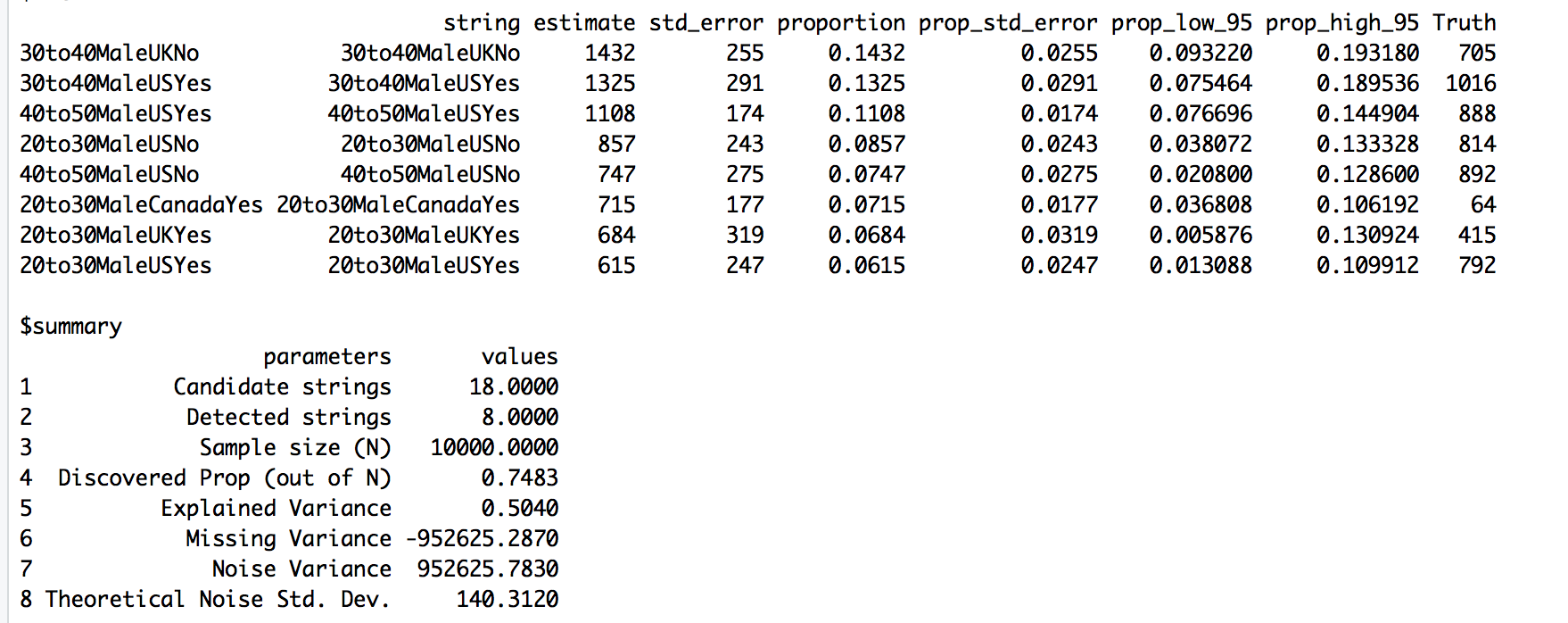


Fig. 4 Truth vs. Estimate

From Fig. 4, it is seen that the first input string (30to40MaleUKNo) is sent to the server 705 times. However, the estimated value by the decode function is 1432. Std error between these values is 255 and the corresponding proportion is 0.0255 which is well below the significance level (alpha = 0.05). It is inferred from the above results that the actual data is privacy preserved and could also effectively used for obtaining statistics.

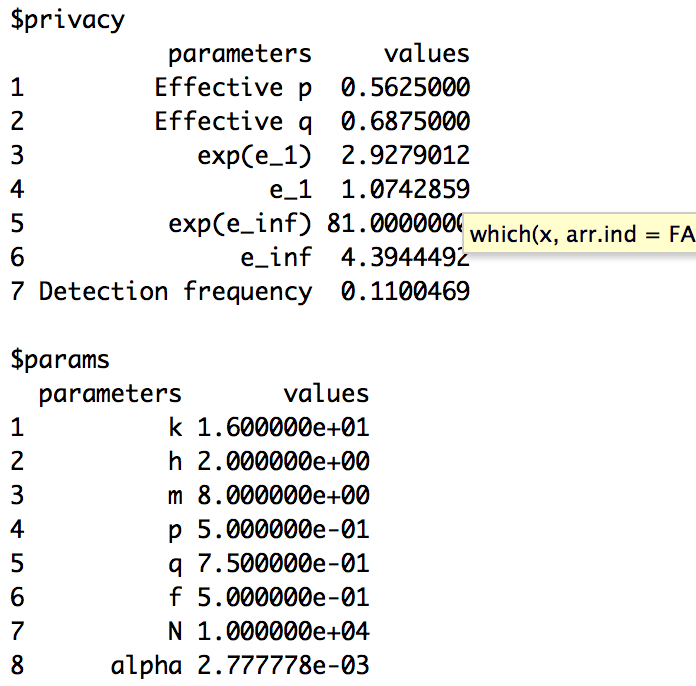


Fig. 5 Privacy parameters used

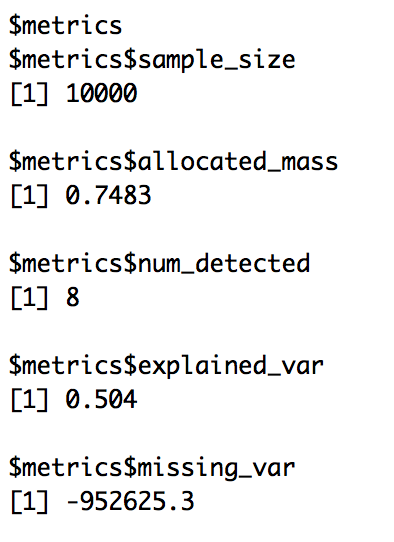


Fig. 6 Results/ metrics

In Fig. 6, the metric $num\_detected represents the number of unique candidate strings obtained after decoding. When the sample size is less, this metric will mostly be zero. This means the server could not decode any survey responses and hence statistical analysis on that data cannot be performed. Therefore, it is often recommended to increase the sample size to few thousands before encoding. For the purpose of this demonstration, I have used 10000 samples. (Refer Fig. 2)

**4.4 Result analysis**

* Input String Explanation 30to40MaleUKNo - 30to40-> Represents the age of the client, Male/Female Represents the gender of the client, UK represents country and Yes/No represent if that user has health issues or not.
* Out of all the data sent to user, truth column says how many values of that particular string are encrypted and sent to the server.
* Estimate values are values decoded from the encrypted values that were sent to the server during encoding.
* From the above results, it is seen that most of the values are decoded with significant accuracy. Also, privacy is preserved because actual bits are not sent to the server.
* String “30to40MaleUKNo”
  + Total number of responses sent to server -> 705 (truth values)
  + Total number of responses found -> 1432 (decoded from noise)
  + Proportion error=0.0255.
  + Comparing the total number of truth values and the estimates, it is seen that estimates are greater than that of actual number of values.
  + The additional number of response (noise) in estimates are added to maintain privacy.
* When the sample size is 0-100, server does not return any response.
* When the sample size is increased, server can decode the inputs and detect the survey responses sent **.**

1. **CONCLUSION**

Results shows that implementing RAPPOR mechanism on health data, improves data privacy and maintains the statistical accuracy. However, technique works well with large number of samples. If sample size is increased from the given data, server can decode the inputs and detect the survey responses sent. With this application of this technique, privacy-preserving statistical analysis of any data is possible even when there is any difference in the statistical parameters between old and new published data.

1. **REFERENCES**

Begum, S., & Nausheen, F. (2018). A comparative analysis of differential privacy vs other privacy mechanisms for Big Data. *2018 2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 512-516). India: IEEE.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of cryptography conference* (pp. 265-284). Berlin, Heidelberg: Springer.

Erlingsson, U., Pihur, V., & Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. *Proceedings of the 21st ACM Conference on Computer and Communications Security* (pp. 1054-1067). Arizona: ACM.

Liu, B., Jiang, Y., Sha, F., & Govindan, R. (2012). Cloud-enabled privacy-preserving collaborative learning for mobile sensing. *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* (pp. 57–70). Canada: ACM.

McSherry, F., & Talwar, K. (2007). Mechanism design via differential privacy. . *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science* (pp. 94–103). Providence, RI, USA: IEEE.