UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
Diciembre 2023, Vol. 16, No. 2, pp. 47–67

Propuesta de estimación de características sensibles utilizando métodos Bayesianos para la Técnica de Conteo de Ítems de Hussain

Proposal of Estimation of Sensible Characteristics using Bayesian Methods for the Hussain Item Counting Technique

Anyi Biviana Moreno Ibagué[a]
anyimoreno@usantotomas.edu.co

Wilmer Dario Pineda Ríos[b]
wilmerpineda@usta.edu.co

Edwin Andrés Cruz Peréz[c]
dec.estadistica@usta.edu.co

Resumen

En este artículo se presenta una propuesta Bayesiana para mejorar el estimador de Hussain en la Técnica de Conteo de Ítems (TCI), con el objetivo de eliminar o reducir la proporción de estimaciones negativas que presenta este estimador cuando se desea estudiar una característica sensible en particular. Se plantea un estimador analítico empleando como distribuciones previas una distribución beta y la distribución uniforme, el análisis se realiza vía simulación planteando diferentes escenarios para el numero de preguntas en el cuestionario, tamaño de la muestra y la proporción conocida de las preguntas no sensibles, obteniendo finalmente la eliminación total de las estimaciones negativas para la proporción de personas que poseen una característica sensible de interés, presentando reducciones significativas en el coeficiente de variación estimado.

***Palabras clave*:** Características Sensibles, Estimador de Hussain, Estimaciones negativas, Técnica de Conteo de Ítems, Técnicas Bayesianas.

Abstract

This paper presents a Bayesian proposal to improve the Hussain estimator in the Item Counting Technique (ICT), with the objective of eliminating or reducing the proportion of negative estimates that this estimator presents when a particular sensitive characteristic is to be studied. An analytical estimator is proposed using

[a]Estudiante Maestría en Estadística Aplicada
[b]Docente Maestría en Estadística Aplicada
[c]Docente Maestría en Estadística Aplicada

as prior distributions a beta distribution and the uniform distribution, the analysis is carried out via simulation by proposing different scenarios for the number of questions in the questionnaire, sample size and the known proportion of non-sensitive questions, obtaining finally the total elimination of the negative estimates for the proportion of people who have a sensitive characteristic of interest, presenting significant reductions in the estimated coefficient of variation.

***Keywords*:** Sensitive Characteristics, Hussain's Estimator, Negative Estimates, Item Counting Technique, Bayesian Techniquess.

1. Introducción

Las estrategias de muestreo son empleadas actualmente en diferentes campos: medicina, economía, sociología, negocios, marketing, tecnologías de la información, psicología, entre otros. Especialmente, en ramas de las ciencias sociales relacionadas con el comportamiento humano, con aspectos que pueden afectar la intimidad de un individuo, las personas no reportan su estado real cuando la característica de interés es de naturaleza sensible, es decir, un comportamiento socialmente desaprobado o temas concernientes a su privacidad. Cuando se pregunta de manera directa sobre la violencia doméstica, aborto, racismo, plagio, evasión de impuestos, infracciones de ley, cultivos ilícitos, consumo de drogas, convicciones religiosas o políticas, problemas de salud, preferencias sexuales, entre otras, las personas usualmente se niegan a responder o mienten en sus respuestas, quizá por miedo, vergüenza, prejuicios sociales, estigmatización social o simplemente por proteger su privacidad.

Según la Ley Estatutaria 1581 de 2012 de Colombia, una persona no está obligada a responder preguntas que atenten a su privacidad e integridad física y/o psicológica, este tipo de preguntas se consideran sensibles o delicadas, los datos sensibles según Pfeiffer (2008), son aquellos que pueden afectar directamente lo más íntimo de una persona, estos datos tienen que ver con enfermedades físicas o psíquicas, conductas sexuales, comportamientos morales, ideologías políticas y/o religiosas. Además, Goffman (1986) que es referenciado por Miric (2005), hace alusión a una característica estigmatizante como "un atributo profundamente desacreditador dentro de una interacción social particular", lo cual puede ser un motivo para inducir el rechazo social. Chaudhuri y Christofides (2013), plantean que es natural que una persona mienta simplemente por no ir en contra de las normas sociales y por ende, estar expuesta a un trato discriminatorio, esto se relaciona con la denominada deseabilidad social, por consiguiente, las respuestas de los encuestados se enmarcan en lo socialmente aceptado (Cobo, 2013), además las interacciones sociales hacen que las personas tiendan a evitar la estigmatización para proteger su propia identidad (Jones, 1984), dando respuestas socialmente deseables a cualquier tipo de preguntas.

El rechazo o la falta de veracidad en las respuestas de los encuestados cuando se hacen preguntas sensibles de manera directa, pueden producir grandes errores

ajenos al muestreo, induciendo tanto un sesgo de rechazo como un sesgo en la respuesta, lo cual influye directamente en la estimación de los parámetros de interés relacionados con la característica que se desea estudiar. Por este motivo, surge la necesidad de plantear alternativas bajo preguntas indirectas, que faciliten la recolección de información sobre variables sensibles, atributos estigmatizantes o temáticas que son socialmente delicadas.

Las técnicas de preguntas indirectas son recomendadas para evitar respuestas evasivas a preguntas sensibles, algunas de las técnicas más usadas son: la Técnica de Respuesta Aleatorizada (TRA), Técnica de Conteo de Ítems (TCI), la Técnica Nominativa y la Técnica de las Tres Cartas. Estas técnicas proponen metodologías en las cuales es prioridad proteger la confidencialidad del individuo y el anonimato de sus respuestas, esto aumenta la cooperación de los encuestados y por ende pueden suministrar respuestas más confiables.

Entre las propuestas de la TCI, se encuentra la planteada por Hussain (2012), la cual emplea solo una muestra para su análisis, estudios recientes demostraron que el estimador insesgado de Hussain presenta proporción de estimaciones negativas, el principal objetivo de esta investigación es proponer una versión mejorada para el estimador de Hussain mediante técnicas de estadística Bayesiana, que permita corregir y/o eliminar la aparición de estimaciones negativas generadas para el estimador de la Técnica de Conteo de Ítems propuesto por Hussain.

Para abordar el desarrollo de este estudio, el presente documento cuenta con la siguiente estructura: la sección 2 muestra el marco teórico y la revisión literaria relacionado con la técnica de conteo de ítems y el estimador de Hussain, en la sección 3 se presenta la propuesta Bayesiana para el estimador de Hussain, desde un enfoque analítico. En la sección 4, se presentan las generalidades de la simulación realizada, en la sección se evidencia los resultados obtenidos para el estimador original de Hussain y el estimador propuesto en términos de la proporción de estimaciones negativas y el coeficiente de variación estimado, por último, la sección 6 refleja las conclusiones obtenidas y posibles recomendaciones para trabajos de investigación futuros.

2. Marco teórico y revisión de literatura

Al realizar una encuesta, posiblemente las personas eluden o mienten ante preguntas específicas, el encuestado no quiere revelar su estado real respecto a un tema que pueda afectar su integridad, busca confidencialidad y privacidad en sus respuestas. Esto se presenta con mayor frecuencia cuando la pregunta es de naturaleza sensible, preguntas que abordan temas de carácter personal, íntimos o delicados, se genera por parte de los encuestados un sesgo de rechazo y un sesgo en la respuesta, que influyen directamente en la estimación de parámetros de interés en una población.

Warner (1965) propuso un método denominado Técnica de Respuesta Aleatorizada (TRA), el cual busca la protección del anonimato del entrevistado con el fin de

reducir la evasión o respuestas falsas a preguntas sensibles, aumentado en índice de respuestas. Esta técnica fue empleada para estimar la proporción de individuos de una población que poseen una característica sensible o característica socialmente rechazada.

De acuerdo a Warner (1965), se supone que un individuo de una población pertenece a un grupo A o un grupo B, grupos mutuamente excluyentes, se quiere obtener la proporción de personas que pertenecen al grupo A, se toma una muestra de tamaño n de la población por medio de un muestreo aleatorio simple con reemplazo, a cada individuo se le proporciona un mecanismo aleatorio, en este caso particular, una ruleta, con secciones marcadas con la letra A con probabilidad p y otras con la letra B con una probabilidad $1-p$, luego el entrevistado gira la ruleta sin que el entrevistador lo observe, debe reportar si la ruleta apunta o no a la letra que representa al grupo que pertenece, el entrevistado requiere decir solamente "si" o "no", si la ruleta apunta o no al grupo correcto, no debe indicar el grupo que señalo la ruleta específicamente.

Esta estrategia tendría como objetivo generar más confianza al entrevistado y por ende una mayor cooperación, ya que el entrevistador no tiene posibilidad de saber con exactitud el grupo al que pertenece el individuo. La estimación se hace mediante máxima verosimilitud, logrando un estimador insesgado, considerando un $p \neq 1/2$.

$$\hat{\pi} = \frac{p-1}{2p-1} + \frac{n^{'}}{(2p-1)n} \tag{1}$$

Donde π es la proporción del grupo A en la población, p es la probabilidad de que la ruleta apunte al grupo A, n tamaño de la muestra y $n^{'}$ el conteo toral de los que reportan "si" y $n - n^{'}$ el conteo de los que reportan "no". En la propuesta planteada se tiene en cuenta la probabilidad de que una persona coopere y el tamaño de muestra necesario para lograr la mayor precisión posible. En los resultados expuestos por Warner (1965), se evidencia que el uso de la TRA logra una mayor cooperación de los encuestados, además de presentar un error cuadrático medio menor en comparación con los métodos de cuestionarios directos. Este método ha tenido variedad de aplicaciones y se han realizado varias mejoras como lo plantean Greenberg et al. (1969a), Mangat and Singh (1990), Van der Heijden et al. (2000), Christofides (2003), entre otros. Esta técnica presenta limitaciones financieras, ya que se requiere un gran tamaño de muestra para obtener intervalos de confianza comparables con los obtenidos con cuestionarios directos, además de emplear más tiempo en aplicar y explicar el procedimiento a los encuestados.

Greenberg et al. (1969b) plantearon una modificación a la TRA propuesta por Warner, con el fin de generar mayor cooperación por parte del encuestado y respuestas más veraces. La técnica consiste en incluir preguntas no relacionadas, de tal manera que el encuestado se enfrente a dos preguntas, una relacionada a la característica sensible y otra que no tenga relación con esta, respondiendo solamente con un "si" o un "No", si se siente representado por alguna de las dos preguntas,

sin expresar directamente a cual de las preguntas atribuye su respuesta, además de considerar dentro de su formulación una probabilidad relacionada a que el encuestado realmente este diciendo la verdad realizaron casos de simulación con el estimador planteado, logrando mejores resultados en términos del error cuadrático medio comparado con el estimador original de Warner.

Mangat and Singh (1990) proponen una alternativa a la técnica de Warner, considerando un procedimiento en dos etapas, comprendiendo un escenario donde los individuos no son completamente sinceros en sus respuestas. El procedimiento planteado implica emplear dos dispositivos aleatorios, sin observación directa del entrevistados, en el primero se hace referencia a dos afirmaciones. 1. Pertenece al grupo sensible, 2. Ir al otro dispositivo aleatorio, cada uno representada con una probabilidad T y $T-1$, respectivamente. En este segundo dispositivo se presenta la posibilidad de pertenecer o no al grupo con cierta característica sensible, con probabilidades iguales a las trabajadas por Warner (1965), p probabilidad de pertenecer al grupo con la característica sensible y $1-p$, no pertenencia al grupo sensible, ambas probabilidades conocidas. El entrevistado usará el segundo dispositivo si así lo indica el resultado del primer dispositivo, al final el encuestado solo debe reportar "si" o "no" según el estado real que posea. Obteniendo un estimador insesgado para la proporción de personas que poseen la característica sensible, mostrando una mayor eficiencia con respecto al estimador de Warner, ya que presenta menor varianza.

Dada la efectividad de la técnica de Warner, Christofides (2003) propone una generalización de la Técnica de Respuesta Aleatorizada (TRA), donde se toma una muestra de tamaño n mediante un muestreo aleatorio simple, la técnica consiste en suministrar a cada individuo un dispositivo aleatorio que contenga L números enteros, donde el entrevistado manifiesta que tan lejos esta del número que arroja el dispositivo aleatorio, lo cual es una diferencia que indica si el individuo tiene o no la característica de interés, eso dependiendo que tan lejos este del valor $L+1$ y 0. Aplicar esta Técnica implicaba limitaciones financieras, ya que se requiere de un tamaño de muestra muy grande para obtener intervalos de confianza comparables con las técnicas de cuestionarios directos, además de implicar una gran cantidad de tiempo para explicar y aplicar la metodología al entrevistado.

La Técnica de Conteo de Ítems (TCI), surge como una alternativa a TRA, fue estudiada inicialmente por Miller (1984), también es conocida como experimento de lista, la técnica consiste en tomar dos submuestras aleatorias independientes, una de control y otra de tratamiento, a la primera se aplica un cuestionario con g ítems no relacionados o preguntas no sensibles, preguntas comunes que no afectan la integridad de la persona y a la segunda submuestra se aplica un cuestionario similar con un ítem adicional que responde a la característica sensible de interés, así la diferencia entre el promedio de respuestas del grupo de control y el grupo de tratamiento, representa una estimación de la proporción de personas en una población que responden de manera afirmativa al ítem relacionado con la variable sensible. Algunos estudios realizados pueden ser referenciados en Droitcour et al. (1991), Chaudhuri y Christofides (2007), Hussain and Shabbir (2010), Imai (2011),

Hussain (2012), Aronow et al. (2015), Blair et al. (2019), Chou et al. (2020).

Hussain and Shabbir (2010) proponen una ingeniosa alternativa a la TCI convencional, en la cual solo es necesario tomar una muestra y por ende no se requiere determinar el tamaño óptimo de las submuestras, esto lo hace más interesante en términos de costo y eficiencia estadística, brindando mejores resultados en las estimaciones de características sensibles garantizando la privacidad del encuestado, el estimador propuesto por Hussain es más eficiente que el estimador de Warner (1965) cuando el número de ítems no relacionados es mayor a tres. Hussain (2012) presenta su propuesta más detallada, mostrando un caso aplicado relacionado con el fraude de exámenes, dicha propuesta es comparada vía simulación con la TCI y TRA, obteniendo un buen desempeño para el caso cuando $\rho_j = \frac{1}{J}$, se realiza la comparación teniendo en cuenta el número de preguntas, tamaño de la muestra, la proporción ρ_j, mostrando resultados más eficientes en términos de la Eficiencia Relativa. Hussain et al. (2013) plantean una mejora desde un punto de vista Bayesiano, proponiendo diferentes distribuciones previas obteniendo estimaciones más precisas en contraste al estimador inicialmente propuesto por máxima verosimilitud.

Trujillo et al. (2020) en su workpaper, plantean extender la TCI de Imai (2011) y la TCI de Hussain (2012) para poblaciones finitas bajo diferentes diseños de muestreo probabilístico, con el fin de estimar el total y a su vez la proporción de individuos que poseen una característica sensible de interés. Presentando el estimador bajo el diseño de muestreo, la varianza y el estimador de la varianza para las dos técnicas, realizando una simulación y aplicación real para los métodos presentados. El resultado de la simulación evidencia la presencia de estimaciones negativas para la proporción de personas que poseen la característica sensible, siendo este un aspecto de especial atención, que en los estudios de Hussain no fue considerado.

El objetivo de esta investigación es proponer una versión mejorada al estimador de Hussain mediante técnicas de estadística Bayesiana, que permita corregir y/o eliminar la aparición de estimaciones negativas generadas para el estimador de la Técnica de Conteo de Ítems propuesto por Hussain cuando se desee calcular la proporción de personas en una población que poseen una característica sensible de interés, aprovechando las ventajas den enfoque bayesiano sobre el frecuentista, en términos de la inclusión de información previa, el manejo de muestras pequeñas, mejor interpretabilidad de resultados, actualización de conocimiento, entre otros.

2.1. Estimador de Hussain

En la técnica de Conteo de Ítem tradicional es necesario utilizar dos muestras, una para el grupo de tratamiento y otra para el grupo control, en 2012 Hussain plantea una nueva alternativa, para la cual solo es necesario tener una sola muestra de tamaño $\boldsymbol{n}$. La técnica consiste en suministrar al entrevistado un cuestionario con J número de ítems. La j-ésima pregunta consiste de un ítem no relacionado y un ítem con la característica sensible, en este caso F_j denota la pregunta no

sensible y H representa la pregunta sensible aplicada al k-ésimo elemento de la muestra. El entrevistado debe responder **1** si él o ella se identifica con al menos una de las dos preguntas, responder **0** en otro caso.

$$\alpha_{kj} = \begin{cases} 1 \text{ si el k-ésimo individuo posee } F_j \text{ o } H \\ 0 \text{ en otro caso} \end{cases}$$

De este modo, se tiene una variable dicotómica α_{kj} de 0 y 1, que se distribuye Bernoulli. Considerando F_j y H como eventos independientes, se determina la probabilidad de la unión de estos eventos como:

$$P(F_j \cup H) = P(F_j) + P(H) - P(F_j \cap H)$$

Considerando las probabilidades de cada evento,

$$P(F_j \cup H) = \rho_j + \pi - \rho_j \pi$$

Por lo tanto,

$$\alpha_{kj} \sim Ber(\rho_j + \pi - \rho_j \pi) \qquad k \in U \qquad j = 1, ..., J \tag{2}$$

Donde ρ_j es la proporción conocida de personas que tienen la característica no sensible en el ítem j, π es la proporción de personas que poseen la característica sensible de interés y U en el conjunto de todos los k individuos de la muestra.

Finalmente, el entrevistado reporta al entrevistador el conteo total de las preguntas con las que se siente identificado, sin especificar un ítem particular.

$$Y_k = \sum_{j=1}^{J} \alpha_{kj} \tag{3}$$

De esta manera, si el entrevistado reporta $Y_k = J$, esto no implica necesariamente que posea la característica sensible de interés, puede que se sienta representado con todos las preguntas no sensibles, asegurando total confidencialidad para el entrevistado. Si $Y_k < J$ implicaría directamente que el encuestado no posee la característica sensible. El valor esperado para la variable respuesta suministrada por el encuestado esta dado por la siguiente expresión:

$$E(Y_k) = \pi \left(J - \sum_{j=1}^{J} \rho_j \right) + \sum_{j=1}^{J} \rho_j \tag{4}$$

Hussain (2012) propone el siguiente estimador insesgado para la proporción de personas que poseen una característica sensible, mediante Máxima Verosimilitud,

$$\hat{\pi}_H = \frac{\bar{Y} - \sum_{j=1}^{J} \rho_j}{J - \sum_{j=1}^{J} \rho_j} \tag{5}$$

Donde

$$\bar{Y} = \frac{\sum_U Y_k}{N} = \frac{\sum_U \sum_{j=1}^{J} \alpha_{kj}}{N}$$

A continuación, se plantea una alternativa a la expresión analítica descrita por Hussain (2012)) para calcular la varianza del estimador. De acuerdo con el estimador de Hussain, ecuación 5, se aplican propiedades distributivas y propiedades de la varianza, obteniendo la siguiente expresión:

$$Var(\hat{\pi}_H) = \frac{Var(Y_k)}{N\left(J - \sum_{j=1}^{J} \rho_j\right)^2}$$

De este modo la varianza del estimador de Hussain depende de la varianza de Y_k, el conteo total que reporta el k-ésimo encuestado, es importante resaltar que las preguntas están relacionadas entre sí, son preguntas dependientes, ya que cada uno de los ítem contiene la pregunta con la característica sensible, por lo tanto:

$$Var(Y_k) = Var\left(\sum_{j=1}^{J} \alpha_{kj}\right) = \sum_{j=1}^{J} Var\left(\alpha_{kj}\right) + \sum \sum_{j \neq j'} Cov\left(\alpha_{kj}, \alpha_{kj'}\right) \tag{6}$$

Desarrollando los cálculos respectivos, se obtienen las siguientes expresiones:

$$\sum_{j=1}^{J} Var(\alpha_{kj}) = \pi(1-\pi)\left(J - 2\sum_{j=1}^{J} \rho_j + \sum_{j=1}^{J} \rho_j^2\right) + (1-\pi)\sum_{j=1}^{J} \rho_j(1-\rho_j)$$

$$\sum \sum_{j \neq j'} Cov(\alpha_{kj}, \alpha_{kj'}) = \pi(1-\pi)\left[J(J-1) - 2(J-1)\sum_{j=1}^{J} \rho_j + \sum \sum_{j \neq j'} \rho_j \rho_{j'}\right]$$

Reemplazando estas dos expresiones en la ecuación 6, se obtiene la varianza de la variable Y_k

$$Var(Y_k) = \pi(1-\pi)\left(J - \sum_{j=1}^{J} \rho_j\right)^2 + (1-\pi)\sum_{j=1}^{J} \rho_j(1-\rho_j)$$

Finalmente, la varianza del estimador de Hussain esta dada por la siguiente expresión:

$$Var(\hat{\pi}_H) = \frac{\pi(1-\pi)}{N} + \frac{(1-\pi)\sum_{j=1}^{J}\rho_j(1-\rho_j)}{N\left(J-\sum_{j=1}^{J}\rho_j\right)^2} \tag{7}$$

3. Estimador de Hussain desde un enfoque Bayesiano

La Estadística Bayesiana ha logrado un gran auge en los últimos años debido a la implementación computacional, en esta se cuantifica la incertidumbre de un evento tomando como punto de partida el estado de información que una persona tiene sobre el tema de interés o parámetro θ. Para reducir la incertidumbre sobre θ se necesita contar con fuentes de información externa e interna a un conjuntos de observaciones $\mathbf{y}$, la primera denominada distribución previa y la segunda la distribución muestral, la mezcla de fuentes de información dan lugar a un estado de información posterior sobre el parámetro de interés, actualizando la información o las creencias que se tenían inicialmente sobre θ, luego de observar los datos reales. La distribución posterior es una función de θ, por tanto la distribución marginal $p(y)$ es una constante, denominada constante de normalización, lo cual implica que la distribución posterior es proporcional al producto entre la verosimilitud y la distribución previa, Hoff (2009) y Gelman et al. (2013).

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{8}$$

El enfoque bayesiano presenta varias ventajas significativas en el análisis estadístico. Permite la incorporación de información previa relevante sobre el tema o fenómeno de estudio, lo que mejora las estimaciones iniciales. Además, facilita la actualización continua del conocimiento sobre el parámetro de interés a medida que se dispone de más datos, cuantificando explícitamente la incertidumbre asociada. Este enfoque también ofrece una interpretabilidad directa de las probabilidades basada en la creencia o confianza en la veracidad de un evento dada la información disponible, a diferencia del enfoque frecuentista, donde la probabilidad se concibe como la frecuencia relativa de un evento en un gran número de repeticiones de un experimento. Otro beneficio importante es su efectividad en el trabajo con muestras pequeñas. A diferencia de la estadística frecuentista, la estadística bayesiana no requiere grandes muestras para realizar inferencias estadísticas robustas y precisas.

3.1. Distribución Muestral - Verosimilitud

3.1.1. Distribución Binomial

Considerando que la respuesta de cada individuo es el conteo total de ítems con los cuales él o ella se siente identificado, se tiene:

$$Y_k = \sum_{j=1}^{J} \alpha_{kj} \tag{9}$$

Bajo el supuesto que $\rho_j = \rho$, la proporción conocida en la población de j-ésimo ítem, es igual para cada uno de las preguntas no relacionadas, por tanto, $\theta = \rho + \pi(1-\rho)$ de este modo la variable respuesta de k-ésimo encuestado Y_k es una variable aleatoria binomial con parámetros J y θ, la cantidad de preguntas del cuestionario y la probabilidad de éxito constante para cada ensayo Bernoulli, respectivamente.

$$Y_k \overset{iid}{\sim} Bin(J, \theta)$$

Considerando que la verosimilitud es igual a la distribución muestral por una constante normalización $\boldsymbol{c}$ y que $\theta = \rho + \pi(1-\rho)$.

$$L(\pi|y) = c \times \prod_{k=1}^{n} \left\{ \binom{J}{y_k} (\rho + \pi - \rho\pi)^{y_k} ((1-\rho) - \pi(1-\rho))^{J-y_k} \right\}$$

Realizando algunos cálculos analíticos y tomando $d = \frac{\rho}{1-\rho}$ se obtiene

$$L(\pi|y) = c \times (1-\rho)^{nJ} \times \prod_{k=1}^{n} \left\{ \binom{J}{y_k} \right\} \sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \pi^m (1-\pi)^{nJ-n\bar{y}} \tag{10}$$

3.2. Distribución Previa

Uno de los beneficios de la estadística Bayesiana es la posibilidad de incluir información a priori sobre el parámetro de interés, mediante distribuciones previas, información que puede ser obtenida de la experiencia de un experto, estudios anteriores, entre otros. La distribución previa representa las creencias iniciales sobre el parámetro de interés antes de observar directamente los datos.

El parámetro de interés π es la proporción poblacional con la característica sensible bajo el estimador de Hussain, por tanto, $0 < \pi < 1$, una de las distribuciones utilizadas para modelar proporciones y sus probabilidades es la distribución beta porque sus valores están restringidos al intervalo [0,1], además permiten una gran flexibilidad para incorporar diferentes tipos de información previa.

$$\pi \sim beta(a, b)$$

Por tanto

$$p(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} \qquad 0 < \pi < 1 \quad a, b > 0 \tag{11}$$

3.3. Modelo beta Binomial

Partiendo de la definición propia del teorema de Bayes, ecuación 8, se plantean dos alternativas para determinar el estimador de la proporción de personas que poseen una característica sensible, la primer propuesta es una formulación analítica y la segunda expresando la posterior como una distribución beta ponderada.

Formulación analítica para el estimador

Aplicando el enfoque Bayesiano, se sustituye la ecuación 10 y 11, en la ecuación 8

$$p(\pi|y) \propto (1-\rho)^{nJ} \times \prod_{k=1}^{n} \left\{ \binom{J}{y_k} \right\} \sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m}\pi^{m}\,(1-\pi)^{nJ-n\bar{y}} . \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}$$

$$p(\pi|y) \propto (1-\rho)^{nJ} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \prod_{k=1}^{n} \left\{ \binom{J}{y_k} \right\} \sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m}\pi^{a+m-1}\,(1-\pi)^{b+nJ-n\bar{y}-1}$$

Ahora se determina la constante de normalización **c**, teniendo en cuenta que $p(\pi|y)$ es una distribución de probabilidad,

$$\mathbf{c} = \left((1-\rho)^{nJ} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \prod_{k=1}^{n} \left\{ \binom{J}{y_k} \right\} \sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \frac{\Gamma(a+m)\Gamma(b+nJ-n\bar{y})}{\Gamma(a+m+b+nJ-n\bar{y})} \right)^{-1}$$

De este modo la distribución posterior para π esta dada por

$$p(\pi|y) = \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m}\pi^{a+m-1}\,(1-\pi)^{b+nJ-n\bar{y}-1}}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \frac{\Gamma(a+m)\Gamma(b+nJ-n\bar{y})}{\Gamma(a+m+b+nJ-n\bar{y})}} \tag{12}$$

Para determinar un estimador Bayesiano para la proporción de personas que poseen una característica sensible de interés, se calcula la media posterior, como el valor esperado de la posterior:

$$E(\pi|y) = \int_{\Theta} \pi p(\pi|y) d\pi$$

Así

$$E(\pi|y) = \int_0^1 \pi \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-k} \pi^{a+m-1} (1-\pi)^{b+nJ-n\bar{y}-1}}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \frac{\Gamma(a+m)\Gamma(b+nJ-n\bar{y})}{\Gamma(a+m+b+nJ-n\bar{y})}} d\pi$$

Finalmente, el estimador desde un enfoque Bayesiano del estimador de Hussain para la proporción de personas que poseen una característica sensible bajo una distribución previa beta esta dado por:

$$\hat{\pi}_{H-BAYES} = \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m+1, b+nJ-n\bar{y})}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m, b+nJ-n\bar{y})} \tag{13}$$

Para determinar la varianza del estimador de Hussain planteado en la ecuación 13 se considera que la a varianza para cualquier variable aleatoria X se puede calcular así

$$Var\left[X\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2 \tag{14}$$

La varianza del estimador de Hussain formulado analíticamente desde un enfoque Bayesiano, esta dada por la siguiente expresión:

$$Var\left[\hat{\pi}_{H-BAYES}\right] = \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m+2, b+nJ-n\bar{y})}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m, b+nJ-n\bar{y})}$$

$$- \left[\frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m+1, b+nJ-n\bar{y})}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(a+m, b+nJ-n\bar{y})}\right]^2$$

3.4. Modelo con previa no informativa

Dado que la distribución previa es el reflejo de las creencias iniciales sobre el parámetro de interés π, se puede presentar la situación de no disponer de información o conocimiento a priori relacionado, en este caso se puede considerar una previa no informativa. Una distribución adecuada para este propósito es la distribución uniforme, a cual representa un conocimiento nulo respecto al parámetro de interés. Esta se puede considerar como un caso especial de la distribución beta, con parámetros fijos $a = 1$ y $b = 1$, el estimador de Hussain Bayesiano estaría dado por la siguiente expresión:

$$\hat{\pi}_{H-BAYES} = \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(m+2, nJ-n\bar{y}+1)}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(1+m, nJ-n\bar{y}+1)}$$

La varianza de este estimador está dada por:

$$Var\left[\hat{\pi}_{H-BAYES}\right] = \frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(m+3, nJ-n\bar{y}+1)}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(m+1, nJ-n\bar{y}+1)}$$

$$- \left[\frac{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(m+2, nJ-n\bar{y}+1)}{\sum_{m=0}^{n\bar{y}} \binom{n\bar{y}}{m} d^{n\bar{y}-m} \mathbf{B}(m+1, nJ-n\bar{y}+1)}\right]^2$$

Un aspecto crucial es que, aunque los estimadores bayesianos pueden no ser insesgados, tienden a tener menor varianza en comparación con los estimadores insesgados tradicionales. Esto se debe a que la incorporación de información previa reduce la incertidumbre en la estimación del parámetro, proporcionando un balance entre el sesgo y la varianza que resulta en un menor error cuadrático medio (MSE).

Además, la capacidad de los métodos bayesianos para manejar situaciones complejas y datos escasos o heterogéneos les da una ventaja significativa en la práctica. La utilización de estructuras jerárquicas y la integración de datos externos permiten una mayor flexibilidad y adaptabilidad en la estimación de parámetros.

En resumen, aunque los métodos bayesianos pueden sacrificar el insesgamiento, la reducción en la variabilidad y la capacidad de integrar información previa justifican su uso, proporcionando estimaciones más estables y precisas en muchos contextos. (Samaniego, 2011)

4. Simulación

En esta sección, se presentan diferentes simulaciones realizadas con el fin de evaluar el desempeño del estimador de Hussain original y la versión planteada bajo el enfoque Bayesiano. Las simulaciones tienen como objetivo comparar la proporción de estimaciones negativas y el coeficiente de variación estimado en diferentes escenarios, variando el tamaño de la muestra, el número de preguntas no sensibles y la proporción de la característica sensible en la población.

4.1. Diseño de la Simulación

Para evaluar el desempeño de los estimadores, se realizaron 10.000 simulaciones utilizando la técnica de remuestreo bootstrap. La población de estudio fue de 1.000 individuos, y se consideraron fracciones de muestreo del 10 %, 20 % y 30 %. La proporción de individuos con la característica sensible (π) varió entre 0.05 y 0.5 con incrementos de 0.05. Además, se incluyeron diferentes números de preguntas no sensibles J en el cuestionario, con valores de 2 a 5. Para cada número de preguntas, se evaluaron tres escenarios de probabilidades conocidas en la población ρ_j, probabilidades altas, medias y bajas, como se estructuró en la tabla 1.

Número de preguntas no sensibles	Proporciones Bajas	Proporciones Medias	Proporciones Altas
2	0.1, 0.1	0.5, 0.5	0.8, 0.8
3	0.1, 0.1, 0.1	0.5, 0.5, 0.5	0.8, 0.8, 0.8
4	0.1, 0.1, 0.1, 0.1	0.5, 0.5, 0.5, 0.5	0.8, 0.8, 0.8, 0.8
5	0.1, 0.1, 0.1 ,0.1, 0.1, 0.1	0.5, 0.5, 0.5, 0.5, 0.5	0.8, 0.8, 0.8, 0.8, 0.8

Tabla 1: Proporciones conocidas en la población para cada ρ_j.

4.2. Métodos Estadísticos y Computacionales

La implementación de las simulaciones se desarrollaron en el software R (versión 4.4.0). En el caso de la versión Bayesiana del estimador de Hussain, para cada uno de los escenarios se generaron datos siguiendo una distribución binomial, donde la probabilidad de éxito se calculó según la proporción conocida de la característica sensible y no sensible en la población. Se utilizó el estimador de Hussain y el planteamiento Bayesiano para calcular las proporciones estimadas de la característica sensible. La varianza y el coeficiente de variación de las estimaciones se calcularon para evaluar la precisión y estabilidad de los estimadores.

5. Resultados

Esta sección presenta los resultados de las simulaciones, comparando el desempeño del estimador de Hussain con su versión mejorada bajo un enfoque Bayesiano. Los resultados se analizan en términos de la proporción de estimaciones negativas y el coeficiente de variación estimado, el tiempo promedio de cómputo para las simulaciones realizadas fue de 1.5 horas.

5.1. Proporción de Estimaciones Negativas

La Figura 1 muestra la proporción de estimaciones negativas para distintos valores de π y diferentes escenarios de probabilidades conocidas ρ_j en la población. Se observa que el estimador de Hussain presenta una proporción significativa de estimaciones negativas, especialmente para muestras pequeñas y altas proporciones conocidas en las preguntas no sensibles. En contraste, el estimador Bayesiano elimina completamente las estimaciones negativas en todos los escenarios evaluados como se muestra en la Figura 2.

Los resultados de la simulación para el estimador de Hussain original, Figura 1, muestran que la proporción de estimaciones negativas se presentan en mayor porcentaje cuando las proporciones ρ_j conocidas en la población toman probabilidades medias y altas, esto para todas la fracciones de muestreo planteadas, siendo $f = 0.1$ y $\pi \leq 0.1$ la que mayores valores representa en las $J = 2, 3, 4, 5$ preguntas no sensibles. También se logra constatar que cuando se tienen preguntas no sensible con alta presencia en la población de estudio, la proporción de estimaciones negativas

dependen del número de preguntas del cuestionario, a mayor número de preguntas menor es la proporción de estimaciones negativas.

El tamaño de la muestra influye en la proporción de estimaciones negativas, bajo todos los casos de proporciones de las preguntas no sensibles, si la muestra es grande la proporción de estimaciones negativas para el estimador de Husssain es menor. Para $J \geq 3$ y $\pi \geq 0.25$ la proporción de estimaciones negativas se hace cada vez más pequeña tendiendo a cero. Para proporciones altas de las preguntas no sensibles en el cuestionario, la proporción de estimaciones negativas puede variar entre el 8 % y 35 %.

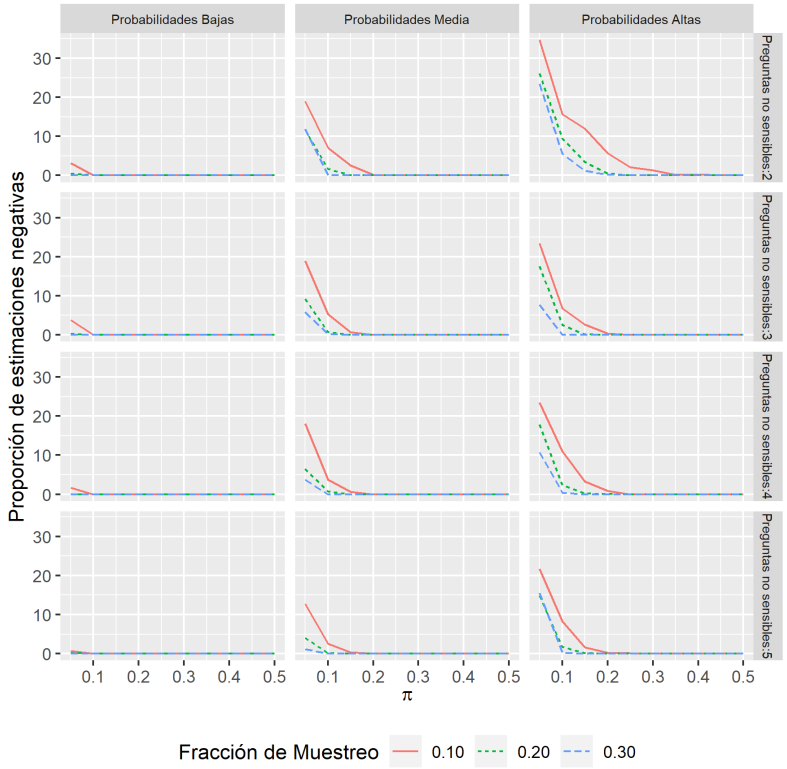

Figura 1: π vs Proporción (%) de estimaciones negativas de la característica sensible de interés, estimador de Hussain 2012

Como resultado principal y quizá el más relevante para esta investigación, es la eliminación total de la proporción de estimaciones negativas para el parámetro π utilizando el enfoque bayesiano, como se muestra en la Figura 2, cumpliendo el objetivo inicialmente propuesto. Se puede observar que para los los diferentes valores de ρ_j, las probabilidades bajas, medias y altas, considerando escenarios de $J = 2, 3, 4$ y 5 preguntas en el cuestionario,la proporción de estimaciones negativas

es nula ante cualquier valor $0.05 \leq \pi \leq 0.5$

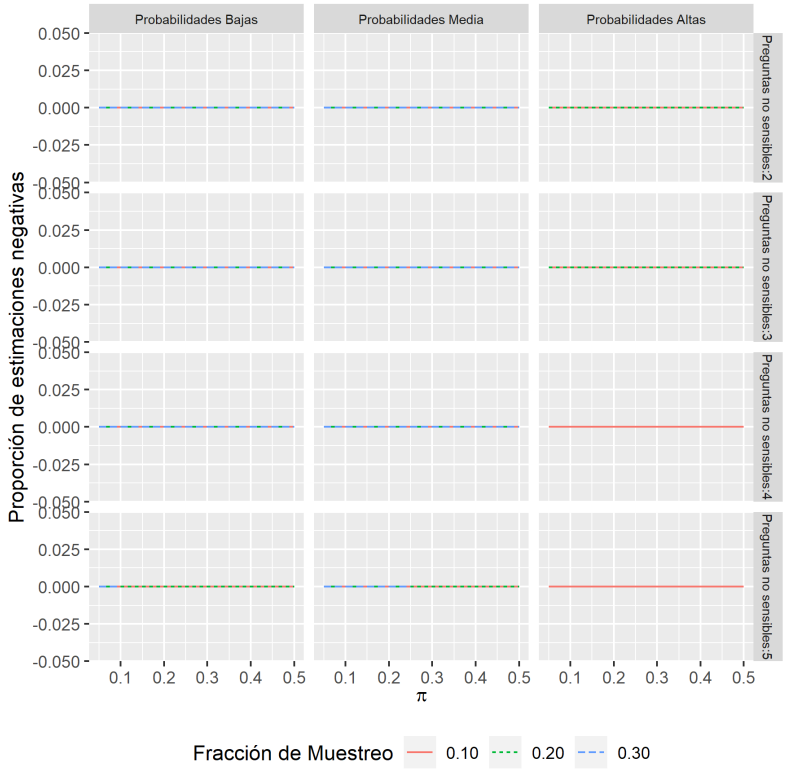

Figura 2: π vs Proporción (%) de estimaciones negativas de la característica sensible de interés, enfoque Bayesiano distribución beta - Binomial, formulación analítica

También se logró identificar que al ser una simulación numérica y dada la forma analítica del estimador, ecuación 13, ante valores grandes de n, el tamaño de la muestra, se presentan valores extremadamente grandes en la combinatoria, al evaluar la función Beta o al realizar el exponente del valor d, no es posible calcular directamente el estimador propuesto, para lo cual se estaría presentando un hallazgo significativo en términos de la precisión numérica y el manejo que se le podría dar a valores numéricos muy grandes. Se aplico la transformación de logaritmo, pero se obtuvo el mismo resultado. Esto se presenta particularmente cuando se considera la fracción de muestre $f = 0.3$, ante probabilidades altas de ρ_j y a medida que el número de preguntas J aumenta.

5.2. Coeficiente de Variación

Para revisar la variabilidad de las estimaciones realizadas mediante la simulación se determina el coeficiente de variación estimado

$$\hat{CV} = \frac{\sqrt{\hat{Var}(\hat{\pi}_H)}}{\hat{\pi}_H} \cdot 100$$

El coeficiente de variación estimado para el estimador propuesto con el enfoque Bayesiano, se presenta en la Figura 3. Se observa que el estimador Bayesiano muestra una reducción significativa en el coeficiente de variación en comparación con el estimador tradicional de Hussain , particularmente cuando se utilizan distribuciones previas informativas. El coeficiente de variación estimado disminuye a medida que el número de preguntas J aumentan y cuando el tamaño de la muestra es mayor.

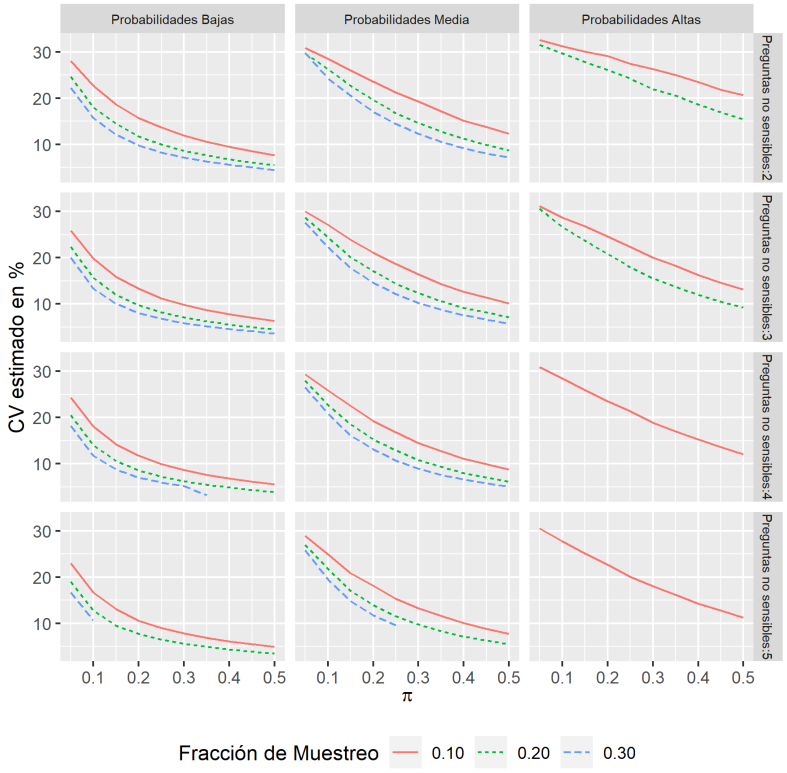

Figura 3: π vs coeficiente de variación estimado (%) enfoque Bayesiano distribución beta - Binomial, formulación analítica

Se obtiene una disminución significativa en el coeficiente de variación estimado em-

pleando el enfoque bayesiano para el estimador de Hussain, logrando coeficientes de variación estimados alrededor del 30 %. Para las probabilidades conocidas ρ_j con valores altas y medias, el coeficiente de variación estimado disminuye a media que el número de preguntas J aumentan y si el valor de π es mayor a 0.2. También se observa que para muestras pequeñas el coeficiente de variación es mayor, particularmente para el caso de probabilidades altas en las preguntas no sensibles el coeficiente de variación estimado varía entre el 10 % y 30 %, mientras que para los casos de probabilidades bajas alcanza valores del 5 %. Para las probabilidades altas en las características no sensibles, para las fracciones de muestreo del 20 % y el 30 % y $J \geq 4$ no se calcula el coeficiente de variación estimado ya que dentro de la simulación realizada se manipulan cantidades muy grandes que R no las calcula con precisión.

6. Conclusiones y Recomendaciones

- Con la implementación de técnicas Bayesianas se logra eliminar la presencia de la proporción de estimaciones negativas para la estimación de la proporción que posee la característica sensible interés en la población, tomando como referencia el estimador de Hussain para la Técnica de Conteo de ítems, utilizando ya sea un a previa beta o una previa no informativa, bajo el supuesto de que las probabilidades de éxito de cada ensayo Bernoulli son iguales, cumpliendo así con el objetivo principal de esta investigación.

- El estimador de Hussain desde el enfoque Bayesiano presenta un mejor comportamiento respecto al coeficiente de variación estimado en relación con el estimador de Hussain normal.

- El hecho de asumir algún conocimiento del parámetro de interés, brinda mejores resultados en términos del coeficiente de variación estimado que al usar una distribución uniforme como distribución previa.

- Conforme a la formulación analítica planteada y la simulación realizada, se presentan retos en la simulación numérica al trabajar con valores muy grandes, particularmente para el caso de las probabilidades altas para las preguntas no sensibles, los valores de ρ_j y a medida que el tamaño de la muestra n aumenta. Al calcular la combinatoria o la función beta, ya sea en la formulación analítica directa o para el calculo de los pesos, estas expresiones arrojan valores muy grandes que no se pueden calcular directamente en R, dando como resultado el no calculo del parámetro de interés.

- Para trabajo futuros se puede profundizar en plantear alternativas que permitan solucionar el reto numérico que se evidencio en el desarrollo de este trabajo, alguna transformación que posibilite la reducción de escala y obtener resultados para diferentes tamaños de muestra y probabilidades altas en la población respeto a las preguntas no sensibles.

- Resulta interesante validar la posibilidad de plantear una estrategia Bayesiana para el estimador de Hussain en poblaciones finitas para cualquier diseño de muestreo.
- Revisar alguna alternativa al considerar diferentes valores de ρ para la construcción de la función de verosimilitud y así hacer un caso aplicable directamente.
- Si los lectores desean revisar las demostraciones realizadas en las expresiones analíticas o parte del código utilizado para generar las simulaciones, se pueden remitir al repositorio del trabajo de grado, el cual pueden encontrar en el siguiente enlace: `https://repository.usta.edu.co/handle/11634/78.`

Recibido: Noviembre 17 de 2023
Aceptado: Diciembre 17 de 2023

Referencias

P. M. Aronow, A. Coppock, F. W. Crawford, and D. P. Green. Combining list experiment and direct question estimates of sensitive behavior prevalence. *Journal of Survey Statistics and Methodology*, 3(1):43–66, 2015.

G. Blair, W. Chou, and K. Imai. List experiments with measurement error. *Political Analysis*, 27(4):455–480, 2019.

T. C. Chaudhuri y Christofides, Arijit y Christofides. *Indirect questioning in sample surveys*. Springer Science & Business Media, 2013.

T. C. Chaudhuri y Christofides, Arijit y s. Item count technique in estimating the proportion of people with a sensitive feature. *Journal of statistical planning and inference*, 137(2):589–593, 2007.

W. Chou, K. Imai, and B. Rosenfeld. Sensitive survey questions with auxiliary information. *Sociological Methods & Research*, 49(2):418–454, 2020.

T. C. Christofides. A generalized randomized response technique. *Metrika*, 57(2): 195–200, 2003.

R. Cobo. Respuesta aleatoria y técnicas de preguntas indirectas. *Departamento de Estadística e Investigación Operativa. Universidad de Granada*, 2013.

J. Droitcour, R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati. The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 1991.

J. Droitcour, R. A. Caspar, M. L. Hubbard, T. L. Parsley, W. Visscher, and T. M. Ezzati. The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, pages 185–210, 2004.

A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL https://books.google.com.co/books?id=ZXL6AQAAQBAJ.

E. Goffman. Stigma: Notes on the management of spoiled identity (a touchstone book), 1986.

A.-L. A. Greenberg, Bernard G y Abul-Ela, W. R. Simmons, and D. G. Horvitz. The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64(326):520–539, 1969a.

B. G. Greenberg, A.-L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz. The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64(326):520–539, 1969b. ISSN 01621459. URL http://www.jstor.org/stable/2283636.

P. D. Hoff. *A first course in Bayesian statistical methods*, volume 580. Springer, 2009.

Y. Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics Data Analysis*, 59:41–51, 03 2013. doi: 10.1016/j.csda.2012.10.006.

Hussain and Shabbir. On item count technique in survey sampling. *Journal of Informatics and Mathematical Sciences*, 2(2 & 3):161–169, 2010.

E. A. y. S. J. Hussain, Zawar y Shah. An alternative item count technique in sensitive surveys. *Revista Colombiana de Estadística*, 35(1):39–54, Jan. 2012. URL https://revistas.unal.edu.co/index.php/estad/article/view/30153.

Z. Hussain, E. ALI SHAH, J. Shabbir, and M. Riaz. On an improved bayesian item count technique using different priors. *Revista Colombiana de Estadística*, 36(2):303–317, 2013.

K. Imai. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494):407–416, 2011. ISSN 01621459. URL http://www.jstor.org/stable/41416378.

E. E. Jones. *Social stigma: The psychology of marked relationships*. WH Freeman, 1984.

Mangat and Singh. An alternative randomized response procedure. *Biometrika*, 77(2):439–442, 1990. ISSN 00063444. URL http://www.jstor.org/stable/2336829.

A. Martinez, M. Borjas, and J. J. Andrade. El fraude académico universitario: el caso de una universidad privada en la ciudad de barranquilla. *Zona Próxima*, (23):1–17, 2015.

J. D. Miller. *A new survey technique for studying deviant behavior*. George Washington University, 1984.

M. Miric. Carga psicosocial del estigma sentido entre las personas que viven con el vih/sida en la república dominicana: Autoestima, depresión y percepción de apoyo social. *Perspectivas psicológicas*, 5:40–48, 2005.

M. L. Pfeiffer. Derecho a la privacidad. protección de los datos sensibles. *Revista Colombiana de Bioética*, 3(1):11–36, 2008.

F. J. Samaniego. *Bayesian vs. Classical Point Estimation: A Comparative Overview*, pages 136–138. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_140. URL https://doi.org/10.1007/978-3-642-04898-2_140.

S. M. Samuels. On the number of successes in independent trials. *The Annals of Mathematical Statistics*, 36(4):1272–1278, 1965. ISSN 00034851. URL http://www.jstor.org/stable/2238127.

C.-E. Särndal, B. Swensson, and J. Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 1992.

C. S. Trujillo, González. Item Count Techniques in Sampling from Finite Population. 2020.

P. G. Van der Heijden, G. Van Gils, J. Bouts, and J. J. Hox. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning: Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, 28(4):505–537, 2000.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. ISSN 01621459. URL http://www.jstor.org/stable/2283137.