**Modelado del Calibre de tubérculos de papa *Solanum phureja* bajo diferentes densidades de siembra mediante regresión binomial negativa cero-inflada**

**Darghan, AE, Bernal, N, Rodríguez, LE**

**Abstract**

A field study was carried out with the cultivation of Creole potato variety *Solanum phureja* to evaluate the influence of the planting density associated to distances between plants of 30,40 and 50 cm and distance between rows and paths of 100 cm on the tuber count of calibers less than 2 cm, 2-4 cm, 4-6 cm and more than 6 cm. The study was conducted at the Marengo Agricultural Center of the National University of Colombia, in the department of Cundinamarca, which has an altitude of 2516 meters above sea level and an average temperature of 14 ° C. According to the characteristics of the climate, the area is classified as Dry Forest Low Mountain. At the time of the harvest of the tubers, they were classified by size and the respective count was made. The modeling of the counts was done by means of usual negative binomial regression and the inflated by zeros option. The zero-inflated negative binomial regression models showed the significant effect of the sowing density on the tuber count in the calibers superior to 4 cm while the negative binomial model showed it in the case of the calibers lower to 4 cm. The results on caliber and density are attributes of interest in both the agronomic management of the crop and in the industrial management of the tubers, so the relationship found can be adopted in both areas to generate the desired attributes of the crop to improve production and guide the process of industrialization.

**Keywords:** PoissonRegression, Negative Binomial Regression, Negative Binomial Regression Zero-inflated, caliber.

**Resumen**

Se realizó un estudio de campo con el cultivo de papa criolla variedad *Solanum phureja* para evaluar la influencia de la densidad de siembra asociada a distancias entre plantas de 30,40 y 50 cm y distancia entre surcos y caminos de 100 cm sobre el conteo de tubérculos de calibres inferiores a 2 cm, de 2-4 cm, 4 – 6 cm y más de 6 cm. El estudio se realizó en el Centro Agropecuario Marengo de la Universidad Nacional de Colombia, en el departamento de Cundinamarca, el cual tiene una altitud de 2516 msnm y una temperatura media de 14 °C. De acuerdo a las características del clima, la zona se clasifica como Bosque Seco Montano Bajo. Al momento de la cosecha de los tubérculos, se hizo su clasificación por calibre y el respectivo conteo. El modelado de los conteos se hizo mediante regresión binomial negativa usual e inflada por ceros. Los modelos de regresión binomial negativa cero-inflada evidenciaron el efecto significativo de la densidad de siembra sobre el conteo de tubérculos en los calibres superiores a 4 cm mientras que el modelo binomial negativo lo evidenció en el caso de los calibres inferiores a 4 cm. Los resultados sobre calibre y densidad son atributos de interés en el manejo tanto agronómico del cultivo como en el manejo industrial de los tubérculos, por lo que la relación encontrada puede adoptarse en ambas áreas para generar los atributos deseados del cultivo para mejorar la producción y guiar el proceso de industrialización.

**Palabras clave:** Regresión Poisson, Regresión Binomial Negativa, Regresión Binomial negativa cero-inflada, calibre.

**Introducción**

La papa criolla Colombia *Solanun tuberosum* Grupo Phureja se encuentra cultivada preferentemente en la parte oriental de los Andes y usualmente entre los 2000 a 3400 msnm. Su distribución geográfica se extiende desde el noroeste de Bolivia, además de gran parte de la región oriental de los Andes Peruanos, hasta Colombia y parte de Venezuela, específicamente en los Andes. Los cultivares pertenecientes a este grupo son precoces, de ahí su nombre nativo en Aymara, phureja, sus tubérculos no tienen periodo de reposo por lo que es posible establecer ciclos de siembra a cosecha tres o cuatro veces al año (Piñeros, 2009). En Colombia, el cultivo de papa se encuentra distribuido en climas fríos con temperaturas de alrededor de 13º C y alturas de 2.000 m.s.n.m., hasta alcanzar zonas de páramo con alturas cercanas a los 3.500 m.s.n.m. y temperaturas de alrededor de 8º C.

Los atributos de calidad de la papa que determinan la aceptación por parte de consumidores no solo se asocian a sus propiedades nutricionales sino también sus características físicas y mecánicas, tales como su dimensión axial, redondez, esfericidad, densidad y porosidad, ángulo de reposo, punto de bioinfluencia, ductilidad, resistencia a la punción de la piel, entre otras. Estos atributos se relacionan con su calibre o diámetro medio, usualmente categorizado por simplicidad en la clasificación, por ejemplo, Buitrago, et al (2004) presenta una clasificación en cuatro categorías según el diámetro, a saber, los grados 0,1,2 y 3, siendo el tamaño más grande la de más de 9 cm (grado 0). Otras clasificaciones que fijan diferentes categorías, a saber, la que aparece en Arias et al, (1996), usan los diámetros de menos de 2cm, de 2-4 cm y la de más de 4 cm y la que se describe en Tabares et al (2009), con categorías de menos de 2cm, de 2-4 cm, de 4-6 cm y más de 6 cm, y finalmente, Escobar y Zaag (1988) los clasificaron en tres diámetros, a saber, menores a 3 cm, entre 3 y 5 cm y más de 5 cm.

Son diversas las investigaciones que relacionan el tamaño y calidad de los tubérculos de papa *solanum phureja* a la densidad de siembra. En la sabana de Bogotá en Colombia se concluyó que en estas condiciones agroclimáticas, las altas densidades de siembra (70 cm entre surcos y 20 cm entre plantas) favorecieron el incremento de tubérculos pequeños (Arias et al, 1996). Los mismos autores relacionaron la densidad de tallos con el número de tubérculos producidos en diferentes densidades de siembra y concluyeron que la densidad de tallos aumentó con un aumento en la densidad de siembra (Bustos et al, 1996).

Engels et al (1993) observaron que la siembra de tubérculos más grandes propiciaba más tempranamente el desarrollo inicial del follaje y que aquellas plantas provenientes de tubérculos de menor calibre en plantaciones de bajas densidades de siembra tendieron a presentar senescencia tardía, además, la cobertura del suelo fue menor en cultivos provenientes de pequeños tubérculos, por lo que concluyeron que la mejor eficiencia en la producción de tallos en plantaciones con tubérculos de menor calibre se compensó con la menor producción de tallos generadores de follaje, lo que se vinculó con una menor capacidad fotosintética.

En un estudio en China subtropical se obtuvo que una mayor densidad de siembra se relaciona con un mayor rendimiento por unidad de área, pero menor rendimiento por planta, evidenciándose consistencia entre las densidades de siembra probadas para todos los ensayos (He et al, 1998).

Wurr et al (2010) encontraron que para producir tubérculos clasificados en un calibre de 4-6.5 cm, o de 4 a 8 cm, la densidad de siembra tuvo un ligero efecto. En zonas de bosque, la productividad de la papa depende principalmente de la densidad de siembra, de hecho, una partición porcentual entre densidad de siembra, nutrición y genotipo, muestran que la mayor contribución se atribuye a la densidad de siembra (por arriba del 55%), con patrones de siembra de 19 y 24 cm de distancia entre plantas y de 70 a 75 cm entre surcos (Vasilyev,2014).

Los diferentes resultados en las investigaciones citadas en este documento y en un poco más de cien adicionales revisados, concluyen basados en análisis de datos usualmente debido a técnicas como el análisis de varianza, análisis de regresión lineal simple y múltiple y en menos casos en regresión no lineal, sin embargo, ninguno de los revisados aunque tratan en muchos de estos el calibre de los tubérculos y evidentemente hacen el conteo de los mismos, usualmente lo asocian al rendimiento y no modelan precisamente los conteos. Resulta interesante reconocer que los conteos pudieran requerir de un ajuste de una distribución específica que tome en cuenta la naturaleza discreta de la variable, a la dispersión de los datos y a la simetría de la distribución sin que se tenga que recurrir a transformaciones estadísticas para encontrar normalidad aproximada de los datos y hacer los análisis usuales antes mencionados. El modelado apropiado puede rendir resultados contrastantes con aquellos análisis donde se ha intervenido en los datos para encontrar los supuestos para la aplicación de una técnica específica, por lo que la elección de una metodología adecuada pudiera rendir resultados diferentes con aquellos donde se ha modificado la distribución inicial, y por ende, la ausencia de significación estadística debida a la densidad de siembra no solo puede atribuirse a los factores involucrados sino al propio modelo utilizado (Nelson y Rawlings,1983;Bennington y Thayne, 1994;Vieckers,2005;Park, et al 2009;Wasserstein y Lazar, 2016)

En ciertas investigaciones de naturaleza agronómica es posible encontrar datos de conteos (unidades que se enumeran) con ceros adicionales a los que se esperarían en aquellas situaciones donde los datos se modelan haciendo uso de la distribución de Poisson, la cual es común en el modelado de conteos con ceros en exceso o inflada por ceros. El problema con el uso de la distribución de Poisson radica en la condición de igualdad en la media y la varianza (equidispersión), por lo que cuanto mayor es el valor de la media mayor será la variabilidad en los datos, y según Hilbe (2014), esto rara vez sucede en datos reales obtenidos en investigaciones de campo, por lo que resulta necesario manipular la sobredispersión (exceso de variabilidad o alta correlación) en el modelo de Poisson o ajustar alguna otra distribución para conteos sin la restricción de equidispersión, como la distribución Binomial Negativa inflada por ceros (Yau et al, 2003). El desconocimiento de este patrón con exceso de ceros puede causar sesgos en la estimación de parámetros o hacer que los investigadores recurran a métodos no paramétricos donde claramente existirán más empates de lo usual, lo que requiere de algún manejo especial de esta condición. Finalmente, se podrán obtener valores de p que conduzcan a conclusiones contradictorias sobre el efecto de factores sobre una variable respuesta, que en nuestro caso es de naturaleza discreta.

En esta investigación se propone el modelado del conteo de tubérculos por calibre (cuatro categorías) mediante modelos de regresión por conteo, específicamente el modelo binomial negativa inflada por ceros usando tres densidades de siembra como factor único en dos repeticiones por densidad y con suficientes unidades experimentales como para evitar problemas en la estimación de los parámetros de sobredispersión debido al tamaño de muestra.

**Materiales y métodos**

**Descripción de la localización del Ensayo**

El estudio se realizó en el Centro agropecuario Marengo de la Universidad Nacional de Colombia, en el departamento de Cundinamarca (74°12'58.51 W; 4°40'52.92 N), el cual tiene una altitud de 2516 msnm, temperatura media de 14 C en un rango de 12C a 18C y precipitación media de 500 a 1000 mm el cual cuenta con un paisaje en planicie fluvio-lacustre y un relieve en terraza lacustre plana (que no excede al 1 %) con suelos son moderadamente profundos y bien drenados . El régimen de humedad es ústico y un nivel freático a menos de 0.5 m del 15 %. De acuerdo a las características de precipitación, temperatura y evapotranspiración, la zona se clasifica como Bosque Seco Montano Bajo.

**Material Vegetal**

El material vegetal utilizado se correspondió al cultivo de papa criolla *Solanum phureja*, utilizando el tubérculo como semilla con el tamaño y forma característica de la especie (tamaño mediano), ojos poco profundos, sin pudrición ni defectos en la piel. Esta variedad con un porte de planta medio y follaje verde claro, con abundante oración, distinguida por su adaptación a días cortos, de origen y distribución en América del Sur, y con centro de diversidad genética al sur de Colombia. Con un desarrollo vegetativo que se da hasta los 35 días después de la siembra (dds), siguiendo la oración hasta los 65 dds, fructificación hasta los 90 dds y finalmente la madurez y senescencia hasta los 120 dds. Esta variedad es precoz (120 días a 2600 msnm), su potencial de rendimiento en condiciones óptimas de cultivo es de 15 a 25 ton.ha−1, sin periodo de reposo y susceptible al virus del amarillamiento de las nervaduras de la hoja (Potato yellow vein virus). Se cultiva en las diferentes regiones de Colombia y en diferentes condiciones de suelo. Es la principal variedad de papa criolla cultivada en el país y hasta la presente es la variedad que se procesa para exportación como precocida congelada (Ñustez, 2011; Rodríguez y Ñustez, 2011).

**Variable respuesta y Explicativa**

A los 120 dds se cosecharon los tubérculos y se contaron según su diámetro en las categorías 2 cm, (2-4] cm, (4-6] cm y > 6 cm. Para estudiar el efecto de la densidad de siembra se fijaron las distancias entre plantas de (30, 40 y 50 cm) todos con separación de 100 cm entre surcos.

**Diseño Experimental**

La siembra se realizó en surcos alineados con precisión según la densidad de siembra, utilizando tres surcos sucesivos según la geometría del lote para cada densidad, con dos repeticiones por densidad, lo que rindió un total de 18 surcos, para un total de 2841 plantas. Aunque la unidad que aportó cada dato fue la planta (tubérculos), la obvia dificultad para aleatorizar una densidad de siembra usando cada planta como unidad experimental, obligó a la aleatorización de las densidades de siembra, cada una con sus tres respectivos surcos (unidad experimental) dentro del lote, registrando los datos de cada planta (unidad de observación). Bajo estas condiciones, el diseño resultó ser una factorial simple en arreglo completamente al azar, tomando las distancias entre plantas como los niveles del factor.

**Análisis Estadístico**

El análisis estadístico involucró inicialmente el componente descriptivo, generando tablas cruzadas para el conteo de tubérculos por densidad y calibre. Adicionalmente, se elaboró una tabla cruzada de promedios, varianzas y cociente entre varianzas y promedio para cada calibre y cada densidad con el objeto de notar la relación posible entre la varianza y la media del conteo de tubérculos, lo cual facilitó la elección de la distribución discreta asociada al proceso de modelado de regresión por conteos y la restricción de sobredispersión. Seguidamente se presentaron los histogramas de conteos por calibre (sin discriminar por densidad) ajustando la distribución de Poisson solo como ilustración de la importancia de explorar los datos para elegir la apropiada distribución de los conteos de tubérculos. El gráfico final incorporó la discriminación por densidad y calibre para los conteos de tubérculos.

En lo que respecta al componente inferencial, se ajustaron diversos modelos del tipo lineal generalizado, presentando finalmente aquellos con los mejores descriptores del ajuste y cumplimiento de supuestos. Los primeros modelos evaluados fueron los de regresión Poisson para conteos, generados uno para cada calibre y utilizando la densidad como factor. Los resultados para el caso Poisson no dieron precisamente los mejores modelos ajustados, no solo por evidenciar sobredispersión, sino porque estos no consideraron el exceso de ceros evidente en la matriz de datos a causa de que en los mayores calibres el número de papas en estos calibres fue nulo en una cantidad de situaciones superior a lo esperado en un modelo de Poisson. Del análisis exploratorio ya era conocido como la distribución Binomial negativa parecía generar el mejor ajuste, por lo que los siguientes modelos de regresión usaron la distribución binomial negativa (sin exceso de ceros) para cada calibre, utilizando como factor la densidad de siembra. Inmediatamente después se obtuvieron los cuatro modelos correspondientes a la opción cero-inflada y se contrastaron mediante el test de Vuong y el criterio de información de Akaike corregido por sesgo, las dos modalidades. Esto se complementó con el gráfico de comparación entre valores observados y estimados de cada modelo en la opción cero-inflada y la convencional, donde se obtuvo el mejor ajuste por calibre.

**Resultados y Discusión**

El tamaño de los tubérculos parece estar regulado por diversos mecanismos y autores como Struik y colaboradores (1990) consideraron la densidad de siembra, el número de tallos por plantas y el número de tubérculos por tallo como variables de importancia en la manipulación y modelado del rendimiento del cultivo. Cotes et al, (2000) recomendaron para esta variedad el calibre de menos de 2cm pero sin efecto de la densidad de siembra, sin embargo, recomendaron una distancia entre plantas de 30 cm con 130 cm entre surcos.

El estudio del calibre juega un papel importante en el propósito de lograr una presentación y conservación de los tubérculos aceptable a consumidores e industria. Investigaciones como la de Ligarreto y Suarez (2003) evidencian la importancia de variables morfométricas tales como los diámetros ecuatoriales y el longitudinal para definir el destino de mercado, industria o consumidores directos, pues las mayores pérdidas en el pelado de las papas se asocian a los calibres más pequeños y a calibres grandes en el caso preparación de hojuelas. En el caso de papa para congelamiento o precocción son deseable los bajos calibres y para el caso de papas fritas o tiras son deseables calibres superiores a los 4 cm. En este sentido, el autor recomendó dimensiones específicas según el destino del producto.

En otro estudio, la disminución de la densidad de siembra dio como resultado un aumento en la cobertura y una reducción de la radiación interceptada acumulada, en la eficiencia en el uso de la luz y en el índice de cosecha, y por ende en los pesos de los tubérculos por unidad de área. La disminución de la densidad de siembra aumentó el número de tubérculos por plántula en todos los calibres, de hecho, una reducción de aproximadamente 146 plantas por m2 a 25 plantas por m2 duplicó el número de tubérculos aprovechables (Van der Veeken y Lommen, 2009).

En este orden de ideas, los resultados que se muestran a continuación se asocian al modelado del conteo de tubérculos de papa criolla para tres diferentes densidades de siembra y para cuatro tamaños de tubérculos.

La Tabla 1 representa el conteo de tubérculos por calibre y el conteo total para cada densidad de siembra, observándose el mayor conteo en las distancias entre plantas de 30 y 40 cm, sin embargo, los porcentajes de tubérculos condicionados por densidad son similares en las tres densidades, lo que a primera vista sugiere que no existen diferencias de importancia en el número de tubérculos obtenidos por densidad.

Tabla 1. **Distribución del conteo de tubérculos por densidad y marca de clase del calibre**

Densidad

Calibres

Subtotal

**≤ 2**

**(2 - 4]**

**(4 - 6]**

**> 6**

d1:(30\*100)

8258(39,5 %)

8337(39,9 %)

4152(19,9 %)

159(0,7%)

20906

d2:(40\*100)

8088(38,3 %)

8538(40,5 %)

4254(20,2 %)

215(1,0%)

21095

d3:(50\*100)

7291(38,4 %)

7064(37,2 %)

4406(23,2 %)

220(1,2%)

18981

Subtotal

23637

23939

12812

594

total(60982)

La Figura 1 ilustra el comportamiento de los conteos en los cuatro calibres sin discriminar por densidad. Sobre cada histograma se superpuso la distribución de Poisson, la que usualmente se usa para el modelado de conteos. Evidentemente los conteos se alejan de la distribución de Poisson esperada, no solo eso, además es claro en los mayores calibres, el exceso de ceros causado por ausencia de papas en este calibre al momento de la cosecha. Por lo que resulta necesario el ajuste de otra distribución por conteos, tal como la Binomial negativa, la que por no tener la misma restricción de sobredispersión que la Poisson, rindió mejor ajuste que el modelo inicial considerado.

Figura 1. **Distribución de conteos de tubérculos por calibre con ajuste de Poisson sin discriminación por densidad de siembra.**

La Tabla 2 presenta los promedios y sus respectivas varianzas de conteos junto al cociente de la varianza y la media para describir la relación que existe entre estas dos estadísticas. En los modelos de conteo, esta relación es relevante para la elección del modelo de probabilidad a elegir para el modelado de los conteos. Los cocientes que se observan en la tabla sugieren que no sería apropiada una elección del modelo Poisson para el ajuste de los datos.

**Tabla 2**. Media (), Varianza (s2) y cociente s2/ por densidad y calibre según densidad.

Calibre(cm)

Densidad: (30\*100)

Densidad: (40\*100)

Densidad:(50\*100)

**s2**

**s2/**

**s2**

**s2/**

**s2**

**s2/**

≤ 2

7,28

22,96

3,2

8,73

27,42

3,1

9,37

38,34

4,1

(2 - 4]

7,35

27,33

3,7

9,22

32,69

3,5

9,08

32,64

3,6

(4 - 6]

3,66

11,74

3,2

4,59

12,86

2,8

5,66

16,01

2,8

> 6

0,14

0,21

1,5

0,23

0,59

2,5

0,28

0,67

2,4

En la Figura 2 se ilustra nuevamente el comportamiento de los conteos, pero ahora discriminados por densidad, para evidenciar los excesos de ceros presentes en los dos calibres superiores para todas las densidades de siembra consideradas. Los calibres se etiquetaron con D1, D3,D5 y D9 para solo indicar el punto medio de los calibres usados. En el caso del calibre etiquetado como D9 se debe a que en este calibre es conocido el punto límite en 6 cm, pero la cota superior se tomó de la literatura donde se muestra que para esta variedad se han tenido calibres hasta los 12 cm, por lo que 9 cm se corresponde con la marca de clase en este calibre-

Figura 2. **Distribución de conteos de tubérculos por densidad de siembra y calibre (D1, D3, D5 y D9)**

El modelado del conteo ha tenido amplia aplicación en el campo de la entomología, pues son diversos los estudios que cuentan el número de insectos para generar medidas de agregación que pudieran relacionarse con umbrales para el manejo del cultivo. Varios estudios incluyen las dos distribuciones más usadas para este tipo de datos, a saber, la Poisson y la Binomial negativa (Cadahia, 1997; Vivas y Notz, 2011). En otras ocasiones, la excesiva aparición de ceros dificulta el uso de estos modelos de probabilidad por lo que se recurre a las opciones cero-infladas o cero-truncadas para tratar con esta situación (Henne, 2012). Precisamente cuando se estudia el calibre en el cultivo de papa como indicador del rendimiento, la naturaleza del dato obliga al uso de las distribuciones más comunes en datos por conteo, a saber, la Poisson y La Binomial negativa y sus modalidades cero infladas o truncadas por ceros, por esta razón, a continuación, se ilustran resultados obtenidos por algunos modelos de prueba y los modelos finalmente propuestos para relacionar el conteo de tubérculos a la densidad de siembra.

El primer modelo ajustado fue el de regresión de Poisson usando como respuesta los conteos de tubérculos solo para el calibre hasta 2 cm y como predictor la densidad de siembra. Este solo se ilustra para justificar el uso de la distribución binomial negativa (Tablas 3 y 4).

Tabla 3. **Modelado por Regresión de Poisson del Calibre hasta 2 cm.**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

1,875

0,017

111,00

2,00e-16

Densidad

0,128

0,008

16,10

2,00e-16

Tabla 4. **Estadísticas de ajuste modelo de Poisson en el calibre D3. (gl: grados de libertad).**

Devianza Nula (gl)

Devianza Residual (gl)

AIC

χ2 (Pearson)

Dispersión

**9556,0(2838)**

9299,2(2837)

13811

9693,6

3,42

En la tabla 4 el estadístico de dispersión es que 3,42 y representa más de lo que se espera en el caso Poisson, y esto corrobora lo que se observó en la tabla 2, donde el cociente fue generalmente superior a 3, indicando sobredispersión (además de una devianza residual superior a los grados de libertad), lo que obligó al uso de otro tipo de distribución, por lo que hasta ahora no se discutió la presencia de un efecto altamente significativo asociado a la densidad de siembra.

En vista de estos resultados ilustrativos previos (algo que usualmente debió hacerse pues es clave entender la razón de selección de un modelo diferente al usual, a saber, el modelo Poisson), pasamos al modelo de regresión Binomial negativo, el que, mediante los gráficos antes mostrados, pareció ser el adecuado en el caso del modelado de los calibres para cada densidad. Se hizo un trabajo por calibres pues fue evidente la presencia de excesivos ceros en los mayores calibres, por lo que, desde ahora, se presentan los resultados del modelado con regresión Binomial negativa y Binomial negativo cero inflada.

En el modelo binomial negativo el parámetro asociado a la media se entiende de la misma manera que el modelo de Poisson, sin embargo, la varianza tiene un alcance mucho más amplio que el permitido por la distribución de Poisson, de hecho la media usa el mismo parámetro que el caso Poisson, pero la varianza se escribe como , con como el parámetro de dispersión. Esta distribución es conocida como binomial negativa del tipo 2 (NB2) por el exponente cuadrático en el segundo término, pero pudiera usarse el tipo NB1, el cual tiene una varianza . Como puede verse, ahora se tiene un modelo de dos parámetros, donde la varianza no está restringida a ser igual a la media, lo que facilitó el modelado, tal como en los datos que estamos evaluando, donde la varianza está resultado un poco más del triple que la media. La binomial negativa como la Poisson permitieron manipular excesos de ceros tal como se encuentran en los diámetros o calibres superiores, por lo que esta distribución está bastante justificada para el modelado de los conteos de tubérculos por densidad. Sin embargo, al igual que el caso de lo hecho en la tabla 3, solo se ilustra el caso del mayor calibre con la binomial negativa, para posteriormente demostrar que los resultados se mejoran con la incorporación de los excesos de ceros.

Tabla 5. **Regresión Binomial Negativa para densidad en el calibre > 6cm**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

-1,965

0,106

-18,52

2,00e-16

Densidad(2)

0,505

0,148

3,40

6,6e-5

Densidad(3)

0,702

0,152

4,62

3,8e-6

Tabla 6. **Estadísticas de ajuste modelo Binomial Negativo en el calibre D9. (gl: grados de libertad).**

Devianza Nula (gl)

Devianza Residual (gl)

AIC

χ2 (Pearson)

Dispersión

**1111,5(2838)**

1088,2(2836)

29,39

0,18

**1111,5(2838)**

Las tablas 5 y 6 muestran la superioridad del modelo binomial negativo sobre el de Poisson, no solo por la disminución en el criterio de información de Akaike, o por la disminución de la devianza nula y residual, de hecho ahora vemos un parámetro de dispersión de 0,18, sin embargo, conocemos el exceso de ceros encontrados especialmente en los dos calibres superiores, por lo que resultó necesario contrastar mediante un test, los modelos binomial negativo con su contraparte que acepta exceso de ceros, por lo que no se discuten exhaustivamente los resultados en las tablas 5 y 6 hasta no haber seleccionado el mejor modelo.

A continuación, se describen los resultados obtenidos mediante el software R para la construcción de las salidas asociadas al proceso de estimación y ajuste de los modelos Binomial Negativo cero-inflado utilizando como factor la densidad, modelando la respuesta por calibre. Las tablas 7 y 8 presentan los resultados del análisis, y se muestran estos dos casos inicialmente ya que son los que no tuvieron un evidente exceso de ceros en el conteo de papas asociados al valor cero. En los dos casos se evidencia el efecto significativo de la densidad de siembra sobre el calibre, tal como se ha encontrado en diversos estudios que evalúan la densidad de siembra como factor que presumiblemente afecta el rendimiento, sin embargo, no podemos aseverar nada con respecto a la similaridad con otros trabajos pues no suele ser este la forma de modelado encontrada en la literatura para los conteos, por lo común del uso del análisis de varianza para evaluar el rendimiento como función de la densidad.

Tabla 7. **Regresión Binomial Negativa cero-inflada para densidad en el calibre ≤ (2cm).**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

1,985

0,019

105,26

2,00e-16

Densidad(2)

1,186

0,028

6,66

2,00e-16

Densidad(3)

0,266

0,029

9,13

2,00e-16

Log(theta)

1,324

0,041

32,07

2,00e-16

Tabla 8. **Regresión Binomial Negativa cero-inflada para densidad en el calibre (2 − 4] cm.**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

2,014

0,020

102,05

2,00e-16

Densidad(2)

0,212

0,028

7,41

1,30e-13

Densidad(3)

0,199

0,030

6,63

3,40e-11

Log(theta)

1,275

0,043

29,94

2,00e-16

La aplicación del test de Vuong permitió la comparación entre los modelos binomial negativo con su contraparte cero-inflada. Este estadístico tiene distribución Normal (0,1). Cuando la significación es menor al 5 % es cuando podemos concluir estadísticamente que la opción cero-inflada es superior (los valores corregidos de AIC y BIC(Bayesiano)) son superiores en la conclusión pues el test es sesgado hacia favorecer la opción cero-inflada). El resultado de la tabla 9 muestra la superioridad de la opción no inflada por ceros en el calibre inferior a 2 cm, y similarmente, la tabla 10 muestra la superioridad del modelo binomial negativo sin exceso de ceros en el calibre de 2-4 cm. Las tablas 11 y 12 muestran que el modelo binomial negativo para modelar el calibre de (2-4] cm fue superior a la opción inflada por ceros y es claro el efecto de la densidad de siembra sobre el conteo de tubérculos de este calibre.

**Tabla 9.** Test de Vuong para comparar la superioridad del modelo cero inflado con si contraparte no inflada por ceros en el calibre inferior a 2 cm.

Estadístico de Vuong

Valor Z

Prob.

Crudo

-1,433

0,076

AIC corregido

-0,589

0,278

BIC corregido

-1,922

0,027

En la tabla 9 no se rechaza la hipótesis nula (dos estadísticos superiores al 5%) que pone como mejor modelo al no inflado por ceros en el calibre menor a 2 cm (p>0.05). La tabla 10 involucra al calibre de 2 a 4 cm, y en este caso tampoco se rechaza la hipótesis nula, poniendo como modelo superior a la opción binomial negativa sin exceso de ceros. Con estos resultados, las tablas 11 y 12 presentan respectivamente las estimaciones del modelo de regresión por conteos en el caso binomial negativo sin exceso de cero y las estadísticas de ajuste para el calibre inferior a los 2 cm.

**Tabla 10.** Test de Vuong para comparar la superioridad del modelo no inflado con su contraparte cero inflada en el calibre inferior a 2-4 cm.

Estadístico de Vuong

Valor Z

Prob.

Crudo

-2,022

0,022

AIC corregido

-1,354

0,088

BIC corregido

-0,633

0,263

Con los resultados hasta este caso, es preciso presentar la tabla del ajuste del modelo binomial negativo sin exceso de ceros para el calibre de 2-4 cm, pues solo se había ilustrado en la tabla 5 el caso de un calibre superior a los 6 cm.

Tabla 11. **Regresión Binomial Negativa para densidad en el calibre menor a 2 cm(D1)**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

1,985

0,019

103,87

2,00e-16

Densidad(2)

0,183

0,028

6,51

3,20e-15

Densidad(3)

0,253

0,029

8,64

2,70e-12

Tabla 12. **Estadísticas de ajuste modelo Binomial Negativo en el calibre D1.**

Devianza Nula (gl)

Devianza Residual (gl)

AIC

Dispersión

**3115(2838)**

3031,6(2836)

16,844

3,61

De forma análoga, en las tablas 13 y 14 se presentan las tablas del ajuste del modelo binomial negativo sin exceso de ceros para el calibre de 2-4 cm:

Tabla 13. **Regresión Binomial Negativa para densidad en el calibre de 2-4 cm.**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

1,994

0,020

101,36

2,00e-16

Densidad(2)

0,227

0,029

7,88

3,20e-15

Densidad(3)

0,212

0,030

7,00

2,70e-12

Tabla 14. **Estadísticas de ajuste modelo Binomial Negativo en el calibre D3..**

Devianza Nula (gl)

Devianza Residual (gl)

AIC

Dispersión

**3157(2838)**

3080(2836)

17042

3,30

La tabla 14 presentael parámetro de dispersión estimado (3,30) , el cual es muy similar a los cocientes obtenidos en la tabla 2. Las tablas 15 y 16 resumen los resultados de la aplicación del test de Voung donde se puso en evidencia la superioridad del modelo binomial negativo con exceso de ceros en los dos calibres superiores (D5 y D9).

**Tabla 15.** Test de Vuong para comparar la superioridad del modelo cero inflado con su contraparte no inflada por ceros en el calibre de 4-6 cm.

Estadístico de Vuong

Valor Z

Prob.

Crudo

6,065

6,61e-10

AIC corregido

5,804

3,23e-9

BIC corregido

5,029

2,46e-7

**Tabla 16.** Test de Vuong para comparar la superioridad del modelo cero inflado con si contraparte no inflada por ceros en el calibre superior a 6 cm.

Estadístico de Vuong

Valor Z

Prob.

Crudo

1,595

0,055

AIC corregido

0,773

0,220

BIC corregido

-1,67

0,047

Tabla 17. **Regresión Binomial Negativa cero inflada para densidad en el calibre de 4-6 cm.**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

1,448

0,026

56,43

2,00e-16

Densidad(2)

0,163

0,036

4,51

6,50e-6

Densidad(3)

0,349

0,037

9,53

2,00e-16

Log(theta)

1,180

0,062

19,14

2,00e-16

Tabla 18. **Estadísticas de ajuste modelo Binomial Negativo cero inflado en calibre D5. (gl: grados de libertad).**

Theta

Log-verosimilitud

gl

**3,253**

-7,18e3

7

Tabla 19. **Regresión Binomial Negativa cero inflada . en el calibre de más de 6 cm.**

Coeficientes

Estimación

Error Estándar

Valor Z

Prob.

Intercepto

-0,870

0,266

-3,26

0,0011

Densidad(2)

0,792

0,219

3,61

0,0003

Densidad(3)

0,641

0,219

2,93

0,0034

Log(theta)

0,217

0,515

0,42

0,6740

Tabla 20. **Estadísticas de ajuste modelo Binomial Negativo cero inflado en calibre D9. .**

Theta

Log-verosimilitud

gl

**1,242**

-1,46e3

7

Los resultados de las tablas 17 y 19 ponen nuevamente en evidencia el efecto significativo de los términos asociados a la densidad de siembra sobre el conteo de tubérculos. En ambas tablas se usó la binomial negativa con función de enlace logarítmico, pues en el caso de enlace logit con la distribución binomial no resultaban significativos los términos en el modelo ajustado, es decir, que el único proceso involucrado es el atribuible a los conteos, ya que un modelo inflado por ceros supone que la presencia de ceros se debe a dos procesos diferentes. Las dos partes de un modelo inflado representan un modelo binario, generalmente un modelo logit para modelar a cuál de los dos procesos se asocia el resultado cero y un modelo de conteo, en este caso, un modelo binomial negativo (el de nuestro caso), para modelar los conteos.

Con todos estos resultados solo queda mostrar gráficamente los resultados del ajuste de los dos primeros calibres en los modelos binomiales negativos y los dos calibres superiores en los modelos binomiales negativos inflados por ceros.

Figura 3. **Comparación de valores observados (Obs.) y predichos (Pred.) (Calibres ≤ (2) y (2 − 4] cm) en el ajuste binomial negativo.**

Figura 4. **Comparación de valores observados (Obs.) y predichos (Pred.) (Calibres (4 − 6] y > 6cm) en el ajuste binomial negativo inflado por ceros.**

Las figuras 3 y 4 corroboran el ajuste obtenido en las modalidades de ajuste binomial negativo y binomial negativo inflado por ceros, siendo muy contrastantes estos gráficos con la figura 1 presentada inicialmente con el ajuste de Poisson. Una vez que el proceso de modelado ha concluido y se percibe claramente lo adecuado del modelado respetando la naturaleza de la variables, puede aseverarse estadísticamente que el factor densidad de siembra explica los conteos para todos los calibres evaluados, lo cual lo corroboran diferentes autores, tales como Escobar y Zaag, (1988), en donde con un aumento de la densidad de siembra de 40000 a 100000 plantas por hectárea incrementó el rendimiento en un 50% pero se generaron tubérculos de menor tamaño, lo que se interpreta como que la densidad tienen efecto sobre el calibre de los tubérculos.

**Conclusiones**

Mediante el uso de modelos adecuados que consideren la naturaleza de las variables se pueden obtener resultados vinculados a estadísticos de prueba que ponen en evidencia efectos que muy seguramente no se detectarían con técnicas convencionales donde se omiten características distribucionales importantes en el proceso de modelado.

Es claro que en los calibres inferiores a 4 cm ajustando modelos binomiales negativos sin exceso de ceros se encontró que los términos asociados a la densidad de siembra resultaron significativos, con lo que puede aseverarse estadísticamente que la densidad de siembra tiene efecto sobre el número de tubérculos.

De forma análoga, en los calibres superiores a 4 cm ajustando modelos binomiales negativos con exceso de ceros se encontró que los términos asociados a la densidad de siembra también resultaron significativos, con lo que puede aseverarse estadísticamente que la densidad de siembra tiene efecto sobre el número de tubérculos en calibres mayores,

El proceso de modelado considera los calibres por separado por tratarse de una respuesta y no de un factor, por lo que el efecto de la densidad es sobre el número de tubérculos condicionados por calibre, pues no se hizo una prueba conjunta de todos los calibres como para aseverar que la densidad afecta el tamaño de los tubérculos, lo cual podría hacerse si se utiliza un promedio ponderado del peso fresco de los tubérculos y se modela ahora una respuesta diferente de naturaleza continua, usando muy seguramente un análisis de varianza multivariante para la matriz de respuestas asociadas a los pesos frescos estimados por calibre y como factor a la densidad de siembra.

**Conclusiones**

**Agradecimientos**

A la Universidad Nacional de Colombia por el apoyo económico para la realización de esta investigación, y al Ingeniero Mateo Cárdenas por el monitoreo complementario del estudio en campo, así como por sus asesorías en el manejo del cultivo y apoyo directo a los autores de la actual publicación.

**Bibliografía**

Arias,V., Bustos,P.,Ñústez,C.E. (1996). Evaluación del Rendimiento en papa criolla (*solanum phureja*) variedad "yema de huevo", bajo diferentes densidades de siembra en la sabana de Bogota. Agronomia Colombiana, Volumen XIII, No.2; pag. 152-161.

Bennington,C.C.,Thayne,W.V. (1994). Use and misuse of mixed model analysis of variance in ecological studies. Ecology, 75(3), pp. 717-722.

Buitrago, G.V.,López,A.P.,Coronado, A.P.,Osorno, F.L. (2004). Determinacion de las características fisicas y propiedades mecanicas de mecanicas de papa cultivada en Colombia. Revista Brasileira de Engenharia Agrícola e Ambiental, v.8, n.1, p.102-110.

Bustos,P., Arias,V.,Ñústez,C.E. (1996). Interrelación entre la densidad de tallos y la tasa de multiplicación de tubérculos en papa criolla (*solanum phureja* juz. et buk) variedad "yema de huevo". Agronomia Colombiana. Volumen XIII. NO.2; pag. 162·168.

Cadhia, D. (1977). Repartición espacial de las poblaciones en Entomología aplicada. Bol. Serv. Plagas, 3: 219-233.

Cotes,J.M., Ñústez,C.E.,Pachón,J.I. (2000). Evaluación de la densidad de siembra y el tamaño del tubérculo semilla en la producción de semilla básica de papa criolla, variedad «yema de huevo» (solanum phureja juz. et buk.). Agronomía Colombiana, 17: 57-57.

Engels,C.,Bedewy, R., Sattelmacher, B. (1993). Effects of weight and planting density of tubers derived from true potato seed on growth and yield of potato crops in Egypt. 1. Sprout growth, field emergence and haulm development. Field Crops Research, 35 ,159-170.

Escobar, V., Zaag, P. (1988). Field performance of potato (*solanum* spp.) cuttings in the warm tropics: influence of planting system, hilling, density and pruning. American Potato Journal, Vol.65, p. 1-10.

He, W.,Struik,P.C.,He,Q., Zhatig,X. (1998). Planting Time and Seed Density Effects on Potato in Subtropical China. J. Agronomy & Crop Science 180, 159-173.

Henne,D.C. (2012). Spatial Patterns and Spread of Potato Zebra Chip Disease in the Texas Panhandle. Plant Disease / Vol. 96 No. 7 ,948-956.

Hilbe, J.M. (2014). Modeling Count Data. Cambridge University Press. New York, USA.

Ligarreto, G.A.,Suarez,M.N. (2003). Evaluación del Potencial de los Recursos Genéticos de papa criolla (*solanum phureja*) por Calidad Industrial. Agronomía Colombiana, 21 (1-2): 83-94.

Nelson,L.A.,Rawlings,J.O.(1983). Ten common misuses of statistics in agronomic research and reporting. Journal of Agronomic Education, Vol. 12, p.100-105.

Ñústez, C. E. 2011. Variedades colombianas de papa. Universidad Nacional de Colombia. Facultad de Agronomía. 48 p.

Park,E. Cho,M.,Ki,C. (2009). Correct Use of Repeated Measures Analysis of Variance. Korean J Lab Med;29:1-9,p.1-9.

Piñeros, C.J. (2009). Recopilación de la investigación del sistema productivo papa criolla. Departamento de Cundinamarca. Federación Colombiana de Productores de Papa-FEDEPAPA. Secretaria de Agricultura y Desarrollo Económico. Colombia.

Rodríguez, L., Ñustez, C., y Estrada, N. (2009). Criolla Latina, Criolla Paisa y Criolla Colombia, nuevos cultivares de papa criolla para el departamento de Antioquia (Colombia). Agronomía Colombiana, 27(3), 289–303.

Tabares,E.D.,Villegas,S.J.,González,L.H.,Cotes, J.M. (2009). Respuesta de la papa (*solanum tuberosum* l.) variedad Diacol Capiro a la Fertilización en un Andisol del oriente Antioqueño, Colombia. Rev. Fac. Nac. Agron. Medellín 62(2): 5099-5110.

Van der Veeken, A.J.H., Lommen. W. J. M. (2009). How Planting Density Affects Number and Yield of Potato Minitubers in a Commercial Glasshouse Production System. Potato Research, 52:105–119.

Vasilyev,A.A. (2014).Dependence of the Yield and Quality of Potato Tubers in the Forest-Steppe Zone of Southern Urals on the Level of Mineral Nutrition and Planting Density. Russian Agricultural Sciences, Vol. 40, No. 6, pp. 422–425.

Vickers,A.J. (2005). Analysis of Variance Is Easily Misapplied in the Analysis of Randomized Trials: A Critique and Discussion of Alternative Statistical Approaches. Psychosomatic Medicine 67:652– 655.

Vivas, L. E. y Notz A. (2011). Distribución espacial en poblaciones de Oebalus insularis Stal (Hemiptera: Pentatomidae) en el cultivo de arroz en Calabozo, estado Guárico, Venezuela. Revista Científica UDO Agrícola 11 (1): 109-125.

Ronald L. Wasserstein & Nicole A. Lazar (2016): The ASA's statement on p-values: context, process, and purpose, The American Statistician, p. 1-16.

Struik, P.C., Haverkort, A. J., Bus, C. B., Dankert, R. (1990). Manipulation of tuber-size distribution of a potato crop. Volume 33, Issue 4, pp 417–432.

Wurr, D.C.E., Fellows,J.R.,Sutherland,R.A.,Allen,J, (1990). Determination of optimum tuber planting density for production of tubers in processing ware grades in the potato variety Record. Journal of Agricultural Science, Cambridge, 114, 11-18.

Yau, K. K. W., Wang, K., Lee, A. H. (2003). Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros.

**Propuesta de estimación del peso fresco por calibre de tubérculos de papa *Solanum phureja* a partir del diámetro medio ponderado y una nueva categorización del calibre.**

**Abstract**

In research in the agricultural sciences it is quite common to find that the shape of agricultural products, as well as the growth of specific structures of a plant or a colony of bacteria is so irregular that it requires the calculation of a weighted average diameter to relate it with other variables of the study. Next, a mathematical methodology of agronomic utility is proposed to obtain the weighted average diameter in potato tubers, where the difficulty of sampling and the large number of observation units complicate the determination of fresh weight by empirical caliber, usually used in the classification of tubers. With the obtaining of the weighted average diameter, an expression to estimate the fresh weight of tubers by size or by any experimental or classification factor was proposed. The linear relationship between the weighted average diameters and the weights per caliber found a low relation in the calibers inferior to 2cm and more than 6 cm in the empirical scale, which can be attributed to the imposed limits of each class (0 and 12 cm). ), so that by means of clusters analysis, a new classification scale of four categories was proposed considering variables such as total fresh weight, plant density, number of tubers and the weighted average diameter itself. The new classification moves the class marks of the extremes a bit, so it is expected that this could improve the estimation of the fresh weights by caliber and its relation with the weighted average diameter. The new scale as well as the expressions for the calculation of the diameter and the fresh weights by caliber are simple to adopt, reason why it is expected that new related works can show their usefulness and their advantage for data analysis.

**Keywords:** average weighted diameter, caliber.

**Resumen**

En las investigaciones en las ciencias agrarias es bastante común encontrar que la forma de los productos agrícolas, así como el crecimiento de estructuras específicas de una planta o de una colonia de bacterias es tan irregular que se requiere del cálculo de un diámetro medio ponderado para relacionarlo con otras variables del estudio. A continuación, se propone una metodología matemática de utilidad agronómica para obtener el diámetro medio ponderado en tubérculos de papa, en donde la dificultad del muestreo y la gran cantidad de unidades de observación complican la determinación del peso fresco por calibre empírico usualmente usado en la clasificación de tubérculos. Con la obtención del diámetro medio ponderado se propuso una expresión para estimar el peso fresco de tubérculos por calibre o por cualquier factor experimental o de clasificación. Se estudió la relación lineal entre los diámetros medios ponderados y los pesos por calibre encontrado baja relación en los calibres inferiores a 2cm y más de 6 cm en la escala empírica, lo que puede atribuirse a los límites impuestos de cada clase (0 y 12 cm), por lo que mediante análisis de clusters se propuso una nueva escala de clasificación de cuatro categorías considerando variables como el peso fresco total, densidad de siembra, número de tubérculos y el propio diámetro medio ponderado. La nueva clasificación mueve un poco las marcas de clase de los extremos por lo que se espera que esto pudiera mejorar la estimación de los pesos frescos por calibre y su relación con el diámetro medio ponderado. La nueva escala así como las expresiones para el cálculo del diámetro y los pesos frescos por calibre son simples de adoptar, por lo que se espera que nuevos trabajos relacionados puedan evidenciar su utilidad y su ventaja para el análisis de datos.

**Palabras clave:** Diámetro ponderado medio, calibre.

**Introducción**

En ciertas investigaciones agronómicas, donde se evalúa el peso fresco de tubérculos, es posible que por simplicidad práctica se haya medido el peso total de los tubérculos para una posterior clasificación por calibre mediante tamizado sin que se haya realizado el pesaje por los respectivos calibres, y que solo se haya registrado el conteo de los tubérculos por calibre y por cualquier otro factor de investigación, por ejemplo, densidad de siembra. Con esta información y a pesar de no poseer el peso para cada calibre, es posible hacer una estimación del peso por calibre condicionado por algún otro factor a partir de la obtención del diámetro medio ponderado (DMP), el cual, como una media ponderada, usa como ponderador el conteo de tubérculos por calibre y la marca de clase asociada al calibre.

Ya que el tamizado de las papas para formar los grupos por calibre solo proporciona una medida indirecta del tamaño de los tubérculos en campo, es posible obtener el diámetro ponderado tal como se ha hecho en la clasificación y tamizado de los agregados del suelo (van Bavel, 1949), o como un porcentaje de los diferentes calibres, que en el caso de partículas del suelo se han clasificado en macroagregados (>2 mm), mesoagregados (250 μm–2 mm), microagregados (53–250 μm) y limo libre + arcilla (<53 μm) (Six et al., 2004), y que en el cultivo de papa criolla se ha hecho en diferentes categorías, por ejemplo, la que aparece en Arias et al, (1996), donde usan los diámetros de menos de 2cm, de 2-4 cm y la de más de 4 cm y la que se describe en Tabares et al (2009), con categorías de menos de 2cm, de 2-4 cm, de 4-6 cm y más de 6 cm, y finalmente, la que describe Escobar y Zaag (1988) , quienes en este caso solo usaron tres diámetros, a saber, menores a 3 cm, entre 3 y 5 cm y más de 5 cm.

Asumamos que una papa o cualquier otro producto agrícola como una naranja o un tomate tienen una forma geométrica regular. Si este es el caso, su forma pudiera describirse usando, por ejemplo, el diámetro, o la longitud de los ejes que forman un elipsoide, sin embargo, como sabemos, los productos agrícolas no poseen formas geométricas perfectamente regulares, por lo que se requerirán múltiples medidas o diámetros para tener una estimación del diámetro medio del producto o de al menos las secciones transversales más importantes, tal como se hace en duraznos, donde se mide el diámetro ecuatorial y el longitudinal para describir la forma de los duraznos en el crecimiento y posteriormente se crea una única medida generada a partir de los diámetros obtenidos separadamente (Quevedo et al, 2017). En el caso de los cocientes de diámetros, el valor reportado es un número que sirve para medir la esfericidad o redondez del fruto. Similarmente, en las manzanas, para obtener el diámetro de las brindillas se miden los diámetros basales y apicales, los diámetros en los sentidos adaxial-abaxial y lateral , para posteriormente calcular la relación de diámetros basales y apicales, así como los diámetros promedio basales y diámetros promedio apicales (Espinoza y Almaguer, 2007). En el caso de el díametro medio ponderado de tubérculos de papa, el procedimiento es similar a lo que se hace en kiwi, donde para cada grado de calidad gustativa, se toma como ponderador la proporción de frutos en dicho grado y se calcula el promedio ponderado o índice TZG (Taste Zespri® Grade) (Godoy et al, 2010)

Aunque este enfoque es simplista y no perfectamente exacto, las formas de los productos agrícolas obtenidos son tales que la suposición esférica en algunos casos no causar problemas serios (uchuvas), sin embargo, algunos autores han reportado mayor variación en frutos cuando estos se refrigeran, lo que evidencia un cambio en su estructura física (García et al, 2010), he ahí la importancia de la evaluación de los diámetros únicos, como promedios aritméticos o promedios ponderados, pues pudieran usarse como indicadores de cambio en la estructura y esto puede servir para el manejo postcosecha de los mismos, en lo que respecta a minimizar el daño.

En el caso del cultivo de papa, algunas investigaciones han mostrado la relación del número de tubérculos con factores como el tipo de semilla utilizada (Morales et a, 2001), entre el peso y número de tubérculos y la densidad de siembra (Cisneros y Herrera,1987; Arismendi, 2002), por lo que relacionar el diámetro medio ponderado para estimar el peso fresco por calibre para un posterior modelado del calibre sea para estimar competición o para modelar los conteos de tubérculos usando como variable explicativa el peso estimado de los tubérculos o el diámetro medio ponderado, es una interesante alternativa en el proceso de modelado de cultivos, más cuando la propuesta usa expresiones de tipo matricial que son manipulables fácilmente en lenguajes de programación como el R.

En la actual investigación, por el manejo de las unidades experimentales donde se cosecharon más de 60 mil tubérculos de papa criolla, ya la tarea de evaluar su calibre fue bastante ardua como para hacer el pesaje de cada tubérculo, por esta razón, fue mucho más simple la clasificación por calibre tal como se descrine en Tabares (2009) y así obtener simplemente el peso total de todos los tubérculos para el calibre discriminando por densidad de siembra para cada planta y no por tubérculo. Sin embargo, como los calibres son variables categóricas, se estableció la marca de cada clase como representante de la categoría, y esta se utilizó como el valor del diámetro inicial y el número de tubérculos por calibre como los ponderadores para generar el diámetro medio ponderado para posteriormente estimar el peso fresco de cada calibre por densidad de siembra, usando los diámetros medios ponderados como nuevos ponderadores del peso total de los tubérculos.

La propuesta es fácil de adoptar e implementar en R, por lo que se espera sea de utilidad a investigadores en las ciencias biológicas especialmente en las ciencias agrarias.

**Materiales y Métodos**

**Descripción del Ensayo**

El estudio se realizó en el Centro agropecuario Marengo de la Universidad Nacional de Colombia, en el departamento de Cundinamarca , el cual tiene una altitud de 2516 msnm y una temperatura media de 14 °C en un rango de 12C a 18C y una precipitación media de 500 a 1000 mm. El material vegetal utilizado se correspondió al cultivo de papa criolla *Solanum phureja*, utilizando el tubérculo como semilla. A los 120 después de la siembra se cosecharon los tubérculos y se contaron según su diámetro en las categorías 2 cm, (2-4] cm, (4-6] cm y > 6 cm. Se utilizaron diferentes densidades de siembra y se fijaron las distancias entre plantas de (30, 40 y 50 cm) todos con separación de 100 cm entre surcos.

La siembra se realizó en surcos alineados por densidad de siembra, con tres surcos sucesivos, con dos repeticiones por densidad, lo que generó un total de 18 surcos, para un total de 2841 plantas. Posterior a la cosecha se registraron los datos de cada planta (unidad de observación) por densidad y repetición y bajo estas condiciones, el diseño resultó ser una factorial simple en arreglo completamente al azar, tomando las distancias entre plantas como los niveles del factor, sin embargo, el propósito de la actual investigación no radicó en la comparación de tratamientos sino en la estimación de los pesos por calibre y por densidad de siembra a partir de los pesos totales de tubérculos por planta.

**Análisis Estadístico**

Partiendo de que el investigador posee los pesos totales de tubérculos (o de cualquier otro producto agrícola) y el número de tubérculos por calibre, el cual pudo haberse discriminado por algún otro factor como por ejemplo la densidad de siembra, el sitio de siembra o alguna otra razón asociada al bloqueo en diseño de experimentos, la metodología implica el uso de algunos operadores utilizados en el contexto de álgebra líneal y algunas otras operaciones definidas en el ámbito computacional. En primer lugar, se obtiene el diámetro medio ponderado (**DMP**), el cual en principio es simplemente un vector para cada grupo de combinación de factores o restricciones de aleatorización, el cual se obtiene mediante:

(1.1)

donde la matriz de dimensión , con como el total de conteos para cada planta en los g calibres conformados, y **c** representa el vector de marcas de clase de cada calibre representado en nuestro caso por **c** = (1, 3, 5, 9), y que en un caso general es un vector de filas con tantas filas como marcas de clase se usen en las clasificaciones de los calibres (dimensión , **c'** representa al vector traspuesto de **c** y como un vector de unos con filas. El cociente entre y se logra haciendo una división elemento a elemento tal como se define es varios lenguajes de programación y que en inglés se suele usar la expresión (elementwise division) o simplemente división elemento a elemento. Con esta simple expresión se obtiene el diámetro medio ponderado como un vector de filas, por lo que se requerirá hacer su separación sea por densidad de siembra o por los otros factores experimentales o de clasificación involucrados en el análisis.

Una vez calculados los diámetros medios ponderados, se usa la expresión que hace la separación o fraccionamiento de la masa por calibre por medio de la expresión:

(1.2)

la cual puede después de una simplificación escribirse como:

(1.3)

siendo diag(**DMP**)−1 la matriz inversa de los diámetros medios ponderados una vez diagonalizados con el operador diag de R, o simplemente usando un proceso de diagonalización encontrada en la literatura. El mismo operador aplica sobre el vector de pesos frescos totales **P,** el cual es de dimensión y sea un vector de unos de filas. Como las tres matrices diagonales previenen de vectores, estas matrices diagonales tienen su inversa garantizada en datos completos (sin celdas faltantes), y en caso de observaciones faltantes, se obtiene su inversa pero se puede cambiar el resultado infinito por cero, tal como se hace en las matrices de pesos espaciales cuando se obtiene el índice de Morán y otros índices (Arbia, 2014).

Para no depender de operadores definidos en lenguajes de programación, la diagonal del vector de **DMP** puede obtener mediante la expresión:

(1.3)

donde es una matriz de dimensión con un 1 en la posición (*i,i*) y ceros en cualquier otra parte; similarmente, es un vector fila de dimensión con un 1 en la posición (1,*i*) y ceros en el resto de posiciones del vector. Si se cambia poro porse obtienen las respectivas matrices diagonales para el peso fresco total y el ponderador o seleccionador. (Banerjee y Roy, 2014).

**Resultados y Discusión**

La figura 1 representa las relaciones lineales encontradas para la partición del por la densidad de siembra. En la figura se representa la línea de regresión lineal simple ajustada por mínimos cuadrados y el coeficiente de correlación obtenido en cada partición, todas resultando en asociaciones lineales evidentes.

**Figura 1.** Diagrama de dispersión del peso fresco y Diámetro medio ponderado sin discriminación por calibre.

Este resultado permite hacer estimaciones a partir del diámetro medio ponderado o estimar el diámetro medio ponderado a partir del peso fresco. Estudios como el de Cardona et al, (2016), han usado el diámetro medio ponderado como indicador al evaluar el efecto de fertilizantes químicos sobre la agregación del suelo. El cambio en el diámetro ponderado se ha usado para evaluar pérdidas en la estructura del suelo asociado a la pérdida de carbono orgánico del suelo debido a prácticas intensivas relacionadas con la preparación de los suelos (Lupi et al, 2007).

Una vez evidenciada la relación entre el DMP y el peso fresco, se obtuvieron con la expresión (1.2) o (1.3) los pesos frescos estimados por calibre y se relacionaron con los propios DMP y se registró la correlación entre el par de variables. La figura 2 discrimina por densidad de siembra el diagrama de puntos entre el par de variables para el calibre inferior a los 2cm, de lo que se observa que las mejores relaciones lineales aparecen en las dos mayores densidades.

**Figura 2.** Diagrama de dispersión del peso fresco estimado para el calibre inferior a 2cm y el Diámetro Medio Ponderado con discriminación por densidad de siembra.

**Figura 3.** Diagrama de dispersión del peso fresco estimado para el calibre de 2 a 4 cm y el Diámetro Medio Ponderado con discriminación por densidad de siembra.

Las figuras 2 y 3 nuevamente evidencian la mejor relación lineal en las dos mayores densidades, con una mejoría en el calibre de (2-4] cm, lo cual en comparación con el calibre inferior a 2 cm puede atribuirse a la marca de clase, pues se usó el 1 como marca cuando se sabe que el cero no es el límite inferior de la clase. Esto podría mejorarse usando alguna forma de optimización donde se modifique la marca de clase a un número superior a 1 pero menor que 2, de modo de establece una mejor marca para así mejorar la estimación del peso fresco por calibre.

Las figuras 4 y 5 muestran los diagramas de dispersión de los dos calibres faltantes. En el caso del calibre de 4 a 6 cm, la correlación entre el par de variables es aceptable y en todas las densidades el valor es similar, sin embargo, en la figura 5 la relación es muy baja, de hecho no sería recomendable usar esta relación entre DMP y peso estimado de tubérculos en este calibre hasta no modificar el diámetro medio ponderado , pues aquí puede existir un problema mayor que en el caso de calibres inferiores a 2 cm, pues superior a 6 cm no posee cota superior, y como fue usada 12 cm por un valor reportado en literatura, es posible que este valor sea muy grande o un caso atípico y no algo común para el calibre, por lo que en este caso si resulta conveniente una disminución del punto 9 cm como marca de clase (valor medio entre 6 y 12 cm) y hacer pruebas con valores inferiores hasta mejorar la estimación.

Sin duda, la relación entre el peso y el DMP es clara, el problema muy seguramente se debe a las marcas de clase en los calibres extremos, pues la cota inferior del calibre menor a 2cm y la cota superior del calibre superior a cm son simplemente medidas supuestas o manejadas por referencia.

**Figura 4.** Diagrama de dispersión del peso fresco estimado para el calibre de 4 a 6 cm y el Diámetro Medio Ponderado con discriminación por densidad de siembra.

**Figura 5.** Diagrama de dispersión del peso fresco estimado para el calibre superior a 6 cm y el Diámetro Medio Ponderado con discriminación por densidad de siembra.

Latabla 1 muestra la distribución de frecuencias y porcentual por calibre y densidad de siembradel conteo de tubérculos, evidenciando un menor conteo en los dos calibres superiores. Esto se muestra con el propósito de crear en el lector el interés de crear una categorización de calibres usando algún método óptimo a alguno de clasificación que maximice las diferencias entre categorías no solo basados en calibres sino en otros criterios, tal como el propio peso fresco, por ejemplo, análisis de clusters.

Usando el método de Ward y las variables diámetro medio ponderado(cm), el número total de tubérculos y el peso fresco total (g)

**Tabla 1**. Distribución de frecuencias y porcentual del número de tubérculos de papa criolla discriminados por densidad de siembra y calibre.

Densidad

**Calibre ≤ 2**

**Calibre (2 - 4]**

**Calibre (4 - 6]**

**Calibre > 6**

**Subtotal**

d1:(30\*100)

8258(39,5 %)

8337(39,9 %)

4152(19,9 %)

159(0,7%)

20906

d2:(40\*100)

8088(38,3 %)

8538(40,5 %)

4254(20,2 %)

215(1,0%)

21095

d3:(50\*100)

7291(38,4 %)

7064(37,2 %)

4406(23,2 %)

220(1,2%)

18981

Subtotal

23637

23939

12812

594

total(60982)

**Figura 6**. Dendrograma para el peso fresco(g), número de tubérculos y DMP marcando los puntos medios de grupo generado.

Estas nuevas marcas de clase pudieran ser de utilidad no solo en el cálculo de un nuevo diámetro ponderado y pesos estimados por calibre, sino evaluar la relación entre peso fresco de calibre y diámetro medio por densidad. Puede notarse que las nuevas marcas están contenidas en la clase que hemos estado desarrollando, pero no coinciden con los puntos medios utilizados. La figura 6 no discrimina por densidad, y como esta es una distancia, el siguiente dendrograma genera unas nuevas marcas usando la misma metodología, pero con la incorporación de la densidad de siembra.

**Figura 7**. Dendrograma para el peso fresco(g), número de tubérculos y DMP marcando los puntos medios de grupo generado y la densidad de siembra.

La incorporación de la densidad baja un poco las marcas de clase, sin embargo, las marcas extremas, es decir, la del calibre inferior a 2 cm y la del calibre superior a 6 cm aunque caen en las categorías empíricas de clasificación del calibre, sus puntos medios se acercan más al límite superior de la clase en el menor calibre y al límite inferior en el calibre mayor.

Una nueva categorización podría generarse a partir de los puntos medios generados en el segundo análisis de clusters, a saber,

**Tabla 2**. Distribución de frecuencias y porcentual del número de tubérculos de papa criolla discriminados por densidad de siembra y calibre según la nueva clasificación

Calibre

**≤ 2.5 cm**

**(2.5 - 4]cm**

**(4 – 5.5]cm**

**> 5.5 cm**

Los valores se aproximaron a una medida práctica por conveniencia y sus límites se generaron con aproximación de los puntos medios de la figura 7.

Aunque esta clasificación puede tener implicaciones en el tamiz y hasta en los destinos industriales que puede tener cada calibre, desde muchos años se ha enseñado el problema de la construcción de escalas empíricas, y una de las razones conocidas se asocia a la paradoja de Simpsom (Selvitella, 2017), por lo que pudieran ser cuestionables algunos análisis de datos donde se involucre el calibre y algún otro factor si se cae en la paradoja, es decir, que dependiendo de la agrupación se obtiene una interpretación diferente de los análisis realizados.

**Conclusiones**

**Referencias**

Arbia,G. (2014). A Primer for Spatial Econometrics with Applications in R. Palgrave Macmillan.UK.

Arias,V., Bustos,P.,Ñústez,C.E. (1996). Evaluación del Rendimiento en papa criolla (*solanum phureja*) variedad "yema de huevo", bajo diferentes densidades de siembra en la sabana de Bogota. Agronomia Colombiana, Volumen XIII, No.2; pag. 152-161.

Arismendi, L.G. (2002). Desarrollo y rendimiento de papa en respuesta a la siembra de semilla–tubérculo inmadura. Revista UDO agrícola 2(1), p. 1-7.

Banerjee,S.,Roy,A. (2014).Linear Algebra and Matrix Analysis for Statistics. CRC Press,Taylor & FrancisGroup. USA.

Cardona,W.A.,Bolaños,M.M.,Chavarriaga,W. (2016). Efecto de fertilizantes químicos y orgánicos sobre la agregación del un suelo cultivado con *Musa acuminata* AA. Acta Agronómica, Vol. 65, Núm. 2.

Cisneros,B.,Herrera,J. (1987). Distancia de siembra y tamaño de tubérculo en la producción de semilla de papa (*solanum tuberosum* l.) en Cartago. Agronomia Costarricense 11(1): p. 65-69.

Escobar, V., Zaag, P. (1988). Field performance of potato (*solanum* spp.) cuttings in the warm tropics: influence of planting system, hilling, density and pruning. American Potato Journal, Vol.65, p. 1-10.

Espinoza,J.R., Almaguer,G. (2007). Estimación de biomasa de órganos aéreos de plantas de manzana cv. agua nueva 2. Revista Chapingo Serie Horticultura 13(2): 179-184.

Garcia,M.C.,Cury,R.K.,Dussán,S.S. (2010). Evaluación poscosecha y estimación de vida útil de guayaba fresca utilizando el modelo de Weibull. Acta Agronómica. 59 (3), p. 347-355

Godoy, C.,Domé,C.,Monti,C. (2010). Determinación de índices de cosecha y calidad en kiwi en el sudeste bonaerense (Argentina). Rev. FCA UNCuyo. Tomo 42. N° 1. p. 53-72.

Lupi, A.M.,Conti,M.,Fernández,R.,Cosentino,D.,López,G. (2007). Efecto de las prácticas de repoblación forestal sobre el carbono orgánico del suelo y la estabilidad de los agregados en el noreste de Argentina. Investigación Agraria: Sistemas y Recursos Forestales, 16(3), 230-240.

Morales,S.D.,Mora,A.R.,Rodríguez,J.E.,Salinas,Y.,Colinas,M.T.,Lozoya,H.(2011). Desarrollo y rendimiento de papa en respuesta a la siembra de semilla–tubérculo inmadura. Rev. Chapingo Ser.Hortic vol.17 no.1,p. 67-75.

Quevedo, E., Darghan, A.E., Fischer,G. (2017). Clasificación de variables morfológicas del duraznero (Prunus persica L. Batsch) ‘Jarillo’ en la montaña santandereana Colombiana mediante análisis discriminante lineal. Revista Colombiana de Ciencias Hortícolas - Vol. 11 - No. 1 - pp. 39-47.

Selvitella,A. (2017). The ubiquity of the Simpson’s Paradox. Journal of Statistical Distributions and Applications . 4:2, p.1-16.

Six, J., Bossut, H., Degryze, S., Denef, K. 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic maƩ er dynamics. Soil Till. Res. 79, 7–31.

Tabares,E.D.,Villegas,S.J.,González,L.H.,Cotes, J.M. (2009). Respuesta de la papa (*solanum tuberosum* l.) variedad Diacol Capiro a la Fertilización en un Andisol del oriente Antioqueño, Colombia. Rev. Fac. Nac. Agron. Medellín 62(2): 5099-5110.

van Bavel, C.H.M. 1949. Mean weight diameter of soil aggregates as a statistical index of aggregation. Soil Sci. Soc. Am. J. 17, 416–418.