E-ISSN: 1115 – 831X P-ISSN: 3027-2963
MPRI-JAVS | INTERNATIONAL JOURNAL OF AGRICULTURAL AND VETERINARY SCIENCE
VOL. 07 NO. 1 - JANUARY, 2025

DETERMINE THE SEX OF PAPAYA AT EARLY STAGE OF THE PLANT USING CATEGORICAL ANALYSIS

OLARINOYE, S.B., AND ZUBBAIR O.R

Statistics and Mathematics Dept, Kwara State Polytechnic, Ilorin

DOI: https://doi.org/10.70382/mejavs.v7i1.027

Keywords:
Papaya, Sex, Logistic, Model, Carica and Regression

Abstract

The research work study papaya plant sex using the two sex categories of male and female. The study tried to fit a model that can be used to detect the sex of the plant early stage using binary logistic regression. A sample size of one hundred and twenty seven was obtained by observation and experiment methods. The variables considered are Colour(x_1), Fruit Production(x2), Leaf branches number(x_3), Stem length before flowering(x_4), Seed Coat Colour(x_5), weight(x_6) and leaf colour(x_7). The Analysis was done using Logistic regression with application of model selection criteria and the variables selected for the model are are Colour(x_1), Leaf branches number(x_3), Stem length before flowering(x_4), Seed Coat Colour(x_5) and the model for the classification is.

$$= \frac{e^{-11.396+41.190x_1-6.410x_3+2.650x_4+0.986x_5}}{1+e^{-11.396+41.190x_1-6.410x_3+2.650x_4+0.986x_5}}$$

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

Introduction

Papaya (*Carica papaya* L.) is an economically important fruit crop that belongs to the Caricaceae family and is native to the Mesoamerica region in America. It is grown worldwide in tropical and subtropical areas. Papaya has a C3 metabolism due to the absence of bundle sheath cells in its leaves. This metabolism makes the plant susceptible to environmental factors, affecting its photosynthetic processes, growth, and fruit development. Light, relative humidity, temperature, and soil composition can impact papaya production. Papaya is a significant fruit crop globally, with more than twelve million tons produced by the top ten producing countries.

Papaya trees (*Carica papaya* L.) are almost entirely propagated from seeds in commercial cultivation because of the economic efficiency. However, this propagation method had some limitation due to the great genetic and phenotypic variability .

As a polygamous species, papaya can have three classes of plants, that is, monoecious, dioecious and monoclinic. Therefore, the crossing within these types, even in controlled situations, will generate seeds with different sex types. In this regard, papaya trees can produce pistillate or female flowers; hermaphrodite; and staminate or male flowers .

Plants which produce pistillate flower are called female and the outcoming fruit are more spherical. On the other hand, plants which bear hermaphrodite flowers produce cylindrical or pyriform fruit with more pulp and they are preferred by the consumer market. Therefore, the sex type determination of papaya trees is commonly done in function to the plant phenotype when they reach the reproductive phase and produce the first flowers. For that, it is necessary to plant three seedlings per planting holes aiming to increase the chance of having at least one hermaphrodite plant . This practice increases the production costs as more seeds and/or seedlings have to be purchased, in addition to maintaining the plants in the field until sexing is performed.

Papaya (Carica papaya L.) is an economically significant plant that produces fruit consumed worldwide due to its organoleptic characteristics. Since their commercial production, papaya fruits have faced several problems, such as pests, which have been partly resolved using transgenic varieties. Nevertheless, a principal challenge in this cultivation is the plant's sex determination. The sex issue in papaya is complex because papaya flowers can

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

bear three sex forms: male, female, and hermaphrodite, which affects their fruit production, shape, and yield. Fruits from hermaphrodite plants are preferred more by consumers than female ones, and male plants rarely produce fruits without commercial value. Chromosomes are responsible for sex determination in papaya, denoted as XY for male, XX for female, and XY^h for hermaphrodite. However, genes related to sex have been reported but are not conclusive. Factors such as the environment, hormones, and genetic and epigenetic background can also affect sex expression. Therefore, in this review, we will discuss recent research on the sex of papaya, from reported genes to date, their biology, and sexing approaches using molecular markers and their advantages.

The sex issue in papaya is complex because papaya flowers can bear three sex forms: male, female, and hermaphrodite, which affects their fruit production, shape, and yield. Fruits from hermaphrodite plants are preferred more by consumers than female ones, and male plants rarely produce fruits without commercial value. Chromosomes are responsible for sex determination in papaya, denoted as XY for male, XX for female, and XY^h for hermaphrodite. However, genes related to sex have been reported but are not conclusive. Factors such as the environment, hormones, and genetic and epigenetic background can also affect sex expression.

Objective (s) of the Study:

- To study the papaya sex types
- To study the contributing variables that determine sex
- To classify the papaya into sex using model
- To test for the fitness of the model

Carica papaya is a herbaceous succulent plant popularly known as pawpaw, and belongs to the Caricaceae family. It is native to the tropics of the Americas but now is widely cultivated in other tropical regions of the world for its edible melon-like fruit, which is available throughout the year . Different parts of the plant are employed in the treatment of different human and veterinary diseases in various parts of the world. For example, in Asian folk medicine, the latex is employed as an abortificient, antiseptic for wound dressing, and as a

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

cure for dyspepsia while in Africa, the root infusion is reputed for treating venereal diseases, piles, and yaws

Carica papaya (CP) is a popular fruit grown in tropical and subtropical countries (Santana et al., 2019). The nutritional and therapeutic qualities of *CP* are widely known. The leaves, among other plant components, have been utilized as therapeutic cures for several ailments (Abdel-Halim et al., 2021). *CP* leaves contain a variety of bioactive components, including phenolic compounds (CGA, ferulic acid, caffeic acid), flavonoids (myricetin and kaempferol), alkaloids (dehydrocarpaine I and II, carpaine, pseudocarpaine), and cyanogenetic chemicals (benzyl glucosinolates). These substances aid in enhancing the antioxidant capacity of blood and lowering the level of lipid peroxidation (*Sachin Kumar Singh and Manoj M. Gadewar* 2023). *Carica papaya* possesses anticancer, anti-inflammatory, antimicrobial, and hypoglycaemic properties.

Carica is basically a tropical plant and is commonly seen in India and Southeast Asian countries like Malaysia and Philippines. The leaves are used for the treatment of fever, pyrexia, diabetes, gonorrhea, syphilis, inflammation, and as dressing for foul wounds. Many reports have also shown that the leaves contain many bioactive agents such as carpaine and nicotinic acid (Bamidele et al., 2008). The anti-inflammatory activity of an ethanolic extract of *Carica papaya* leaves was validated in rats in the standard carrageenan-induced paw edema, cotton pellet granuloma, and formaldehyde-induced arthritis models (*Arora and Baliga 2013*). Ethanolic extract is shown to contain saponins, cardiac glycosides, anthraquinones, and reducing sugars in addition to flavonoids, alkaloids, and tannins. Alkaloids, flavonoids, saponins, tannins, and glycosides, which possess anti-inflammatory properties, are mainly responsible for the plant's beneficial effects.

The papaya (*Carica papaya* L.) belongs to the family *Caricaceae*. It is often called ‘pawpaw’ in Australia and ‘tree melon’ in some other countries, but it is different from the North American ‘pawpaw’ (*Asimina triloba* Dunal), which is a member of the family *Annonaceae. Carica papaya* is believed to be native to tropical America, its region of origin perhaps being southern Mexico and neighbouring Central America. Spanish explorers took it to the Caribbean and South East Asia in the sixteenth century. The large number of seeds in the fruit and their long viability are among the factors that are thought to have

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

contributed to the wide geographical distribution of the fruit (*Singh and Sudhakar Rao 2011*).

The papaya (*Carica papaya* L.) belongs to the family *Caricaceae*. It is often called 'pawpaw' in Australia and 'tree melon' in some other countries, but it is different from the North American 'pawpaw' (*Asimina triloba* Dunal), which is a member of the family *Annonaceae. Carica papaya* is believed to be native to tropical America, its region of origin perhaps being southern Mexico and neighbouring Central America. Spanish explorers took it to the Caribbean and South East Asia in the sixteenth century. The large number of seeds in the fruit and their long viability are among the factors that are thought to have contributed to the wide geographical distribution of the fruit (*Singh, S.P. and Sudhakar Rao, D.V.* 2011).

Methodology

Sample Size

The sample size estimation will be obtained through the use of choice of population (percentage). The guide of error will also be at 5% the sample estimate will be obtain through the below

$$P \pm 2\delta \text{ ----------1}$$

$$2\delta = 2\sqrt{\frac{pq}{n}} = 5 \text{ ----------2}$$

$$n = \frac{4pq}{25}$$

Where P = percentage sample size used

Method of Analysis

Logistic Regression

Logistic regression is the appropriate analysis to conduct when the dependent variable is dichotomous(binary). Like all regression analysis, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

Binary Logistic Regression Major Assumptions

- The independent variable should be dichotomous in nature (e.g., presence vs. absent).
- There should be no outliers in the data, which can be assessed by converting the continuous predictors to standardized scores and removing values below -3.29 or greater than 3.29.
- There should be no high correlations(multicollinearity) among the predictors. This can be assessed by a correlation matrix among the predictors. Tabachnick and Fidell(2013), suggest that as long correlation coefficients among independent variables are less than 0.90, the assumption is met.

At the center of the logistic regression analysis is the task estimating the log odds of an event. Mathematically, logistic regression estimates a multiple linear regression function defined as:

$$\text{Logit(p)} = \log \frac{p(y=1)}{1-(p=1)} = \beta_0 + \beta_1 - x_2 + \beta_2 - x_2 + \ldots + \beta_p - x_m$$

For i = 1, 2,… , n

Types of Logistic Regression are:

Binary logistic regression is the statistical technique used to predict the relationship between the dependent variable(Y) and the independent variable(X), where the dependent variable is binary in nature. For example, the output can be Success/Failure, 0/1, True/False or Yes/No. This is the type of logistic regression that we've been Focusing on in this post.

Multinomial logistic regression is used when you have one categorical dependent variable with two or more unordered levels (i.e., two or more discrete outcomes). It is very similar to logistic regression except that here, you can have more than two possible outcomes. For t, let's imagine that you want to predict what will be the most-used transportation type in the year 2030. The transport type will be the dependent variable, with possible outputs o train, bus, tram and bike (for example).

Ordinal logistic regression is used when the dependent variable (Y) is ordered (i.e., ordinal). The dependent variable has a meaningful order and more than

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

two categories or levels. Examples of such variables might be t-shirt size (XS/S/M/L/XL), answers on an opinion poll (Agree/Disagree/Neutral), or scores on a test (Poor/Average/Good).

The relationship between the predictor and response variables is not a linear function in logistic regression function is used, which is the logit transformation of θ:

$$\theta = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i)}}{1+e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i)}}$$

Where α = the constant of the equation and, β = the coefficient of the predictor variables. An alternative form of the logistic regression equation is:

$$\text{Logit}\,[\theta(x)] = \log[\frac{\theta(x)}{1-\theta(x)}] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i$$

Wald Test

A Wald test is used o test the statistical significance of each coefficient (β) in the model. A Wald test calculates a Z-statistic, which is:

$$z = \frac{\hat{\mathrm{B}}}{SE}$$

Likelihood-Ratio Test

The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the full model (L1) over the maximized value of the likelihood function for the simpler model (L0).

The likelihood-ratio test statistic =

$$-2\log\frac{L_0}{L_1} = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)$$

B1≠ 0 against the null hypothesis that β1 = 0

Data Presenatation

sn	Grp (y)	Colour (xl)	fruit prodn (x2)	no. leaf branches (x3)	stem lenth before flowering(x4)	Seed coat colour(x5)	Weight (x6)	leaf colour(x7)
1	1	1	1	9	2.4	1	1	1

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

2	1	1	1	9	1.9	1	1	1
3	1	1	1	9	2.1	2	2	2
4	1	1	1	12	2.2	1	1	1
5	2	2	2	9	1.9	1	1	1
6	1	1	1	9	2.3	1	1	1
7	1	1	1	9	2.1	1	1	1
8	2	2	2	9	2	2	2	2
9	1	1	1	9	2.1	1	1	1
10	1	1	1	9	2.3	1	1	1
11	1	1	1	9	2.1	1	1	1
12	2	2	2	12	4.2	2	2	2
13	2	2	2	12	2.8	2	2	2
14	2	2	2	12	2.6	2	2	2
15	1	1	1	12	2.7	2	2	2
16	2	2	2	12	2.7	1	1	1
17	2	2	2	12	2.9	2	2	2
18	2	2	2	9	2.7	2	2	2
19	2	2	2	12	3	2	2	2
20	1	1	1	12	2.7	2	2	2
21	2	2	2	12	4.1	1	1	1
22	2	2	2	12	2.7	2	2	2
23	2	2	2	12	3	2	2	2
24	1	1	1	12	2.6	2	2	2
25	2	2	2	12	2.7	2	2	2
26	1	1	1	9	2.2	1	1	1
27	1	1	1	9	1.9	1	1	1
28	1	1	1	9	2.3	2	2	2
29	2	2	2	9	2.1	1	1	1
30	1	1	1	9	2	2	2	2
31	1	1	1	9	2.1	1	1	1
32	2	2	2	9	2.3	2	2	2
33	1	1	1	9	2.1	1	1	1
34	2	2	2	12	4.2	2	2	2
35	2	2	2	12	2.8	2	2	2
36	1	1	1	12	2.6	1	1	1
37	2	2	2	12	2.7	2	2	2
38	2	2	2	12	2.7	2	2	2
39	2	2	2	12	2.9	2	2	2
40	1	1	1	9	2.7	1	1	1

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

41	2	2	2	12	3	2	2	2
42	1	1	1	9	2.1	1	1	1
43	1	1	1	9	2	1	1	1
44	2	2	2	9	2.1	2	2	2
45	1	1	1	9	2.3	1	1	1
46	1	1	1	9	2.1	1	1	1
47	2	2	2	12	4.2	2	2	2
48	2	2	2	12	2.8	2	2	2
49	2	2	2	12	2.6	2	2	2
50	2	2	2	12	2.7	1	1	1
51	1	1	1	12	2.7	2	2	2
52	2	2	2	12	2.9	2	2	2
53	2	2	2	9	2.7	1	1	1
54	2	2	2	12	3	2	2	2
55	1	1	1	12	2.7	1	1	1
56	2	2	2	12	4.1	2	2	2
57	1	1	1	12	2.7	1	1	1
58	2	2	2	12	3	2	2	2
59	1	1	1	9	2.2	1	1	1
60	1	1	1	9	1.9	1	1	1
61	2	2	2	9	2.3	2	2	2
62	1	1	1	9	2.1	1	1	1
63	1	1	1	9	2	1	1	1
64	2	2	2	9	2.1	1	1	1
65	1	1	1	9	2.3	2	2	2
66	1	1	1	9	2.1	1	1	1
67	2	2	2	12	4.2	2	2	2
68	2	2	2	12	2.8	1	1	1
69	1	1	1	12	2.6	2	2	2
70	2	2	2	12	2.7	1	1	1
71	1	1	1	12	2.7	2	2	2
72	2	2	2	12	2.9	1	1	1
73	1	1	1	9	2.7	2	2	2
74	2	2	2	12	3	2	2	2
75	1	1	1	12	2.7	1	1	1
76	2	2	2	12	4.1	2	2	2
77	2	2	2	12	2.7	2	2	2
78	1	1	1	12	3	2	2	2
79	2	2	2	12	2.6	1	1	1

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

80	2	2	2	12	2.7	2	2	2
81	1	1	1	9	2.2	1	1	1
82	1	1	1	9	1.9	1	1	1
83	2	2	2	12	3	2	2	2
84	2	2	2	12	2.7	2	2	2
85	2	2	2	12	4.1	1	1	1
86	1	1	1	12	2.7	2	2	2
87	2	2	2	12	3	2	2	2
88	2	2	2	12	2.6	1	1	1
89	2	2	2	12	2.7	2	2	2
90	1	1	1	9	2.2	1	1	1
91	2	2	2	9	1.9	1	1	1
92	1	1	1	9	2.3	2	2	2
93	1	1	1	9	2.1	1	1	1
94	2	2	2	9	2	1	1	1
95	1	1	1	9	2.1	2	2	2
96	1	1	1	9	2.3	1	1	1
97	1	1	1	9	2.1	1	1	1
98	2	2	2	12	4.2	2	2	2
99	2	2	2	12	2.8	2	2	2
100	1	1	1	12	2.6	1	1	1
101	2	2	2	12	2.7	2	2	2
102	2	2	2	12	2.7	2	2	2
103	1	1	1	12	2.9	1	1	1
104	2	2	2	9	2.7	2	2	2
105	1	1	1	9	2.1	1	1	1
106	1	1	1	9	2.2	2	2	2
107	2	2	2	9	1.9	1	1	1
108	1	1	1	9	2.3	1	1	1
109	1	1	1	9	2.1	2	2	2
110	2	2	2	9	2	1	1	1
111	1	1	1	9	2.1	1	1	1
112	2	2	2	9	2.3	2	2	2
113	1	1	1	9	2.1	1	1	1
114	2	2	2	12	4.2	2	2	2
115	2	2	2	12	2.8	2	2	2
116	1	2	2	12	2.6	2	2	2
117	2	2	2	12	2.7	1	1	1
118	2	2	2	12	2.7	2	2	2

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]

119	1	2	2	12	2.9	1	1	1
120	2	2	2	9	2.7	2	2	2
121	1	1	1	12	3	2	2	2
122	2	2	2	12	2.7	2	2	2
123	2	2	2	12	4.1	1	1	1
124	1	1	1	12	2.7	2	2	2
125	2	2	2	12	3	1	1	1
126	1	1	1	12	2.6	2	2	2
127	2	2	2	12	2.7	1	1	1

Analysis

Classification Table[a]

	Observed		Predicted		
			y		Percentage Correct
			1.00	2.00	
Step 1	y	1.00	58	2	96.7
		2.00	0	67	100.0
	Overall Percentage				98.4

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	x1	41.190	7696.132	.000	1	.996	7.738E17
	x3	-6.410	1992.939	.000	1	.997	.002
	x4	2.650	4.291	.381	1	.537	14.150
	x5	.986	1.471	.449	1	.503	2.680
	Constant	-11.396	15392.262	.000	1	.999	.000

a. Variable(s) entered on step 1: x1, x3, x4, x5.

Interpretation of Results

The Binary Regression using the model selection criteria with seven variables x1,x2,x3,x4,x5,x6,x7. Three of the variables were dropped for the equation and the variables fitted are Colour (x1), No. of leaf branch (x3), stem length before flowering(x4) and Seed Colour(x5) .

The Logistic Regression Model is

$$= \frac{e^{-11.396+41.190x_1-6.410x_3+2.650x_4+0.986x_5}}{1+e^{-11.396+41.190x_1-6.410x_3+2.650x_4+0.986x_5}}$$

Reference

Adejuwon Adewale Adeneye(2014), Journal in Toxicological Survey of African Medicinal Plants, 23-41, 2014

Arora .R and Baliga M.S.(2013), Journal in Bioactive Food as Dietary Interventions for Arthritis and Related Inflammatory Diseases, 2013

Harrell, Frank E. (2015). Regression Modeling Strategies. Springer Series in Statistics (2nd ed.). New York; Springer. doi:10.1007/978-3-319-19425-7. ISBN 978-3-319-19424-0.

Howard M. Taylor &Samuel K (1974): First Course in Stochastic Process. Academic press. Inc London

Hosmer, David W.; Lemeshow, Stanley (2000). Applied Logistic Regression (2nd ed.). Wiley. ISBN 978-0-471-35632-5 Cramer 2002, p. 10–11.

Palei, S. K.; Das, S. K. (2009). "Logistic regression model for prediction of roof fall risks in bord and pillar workings in coal mines: An approach". Safety Science. **47**: 88–96. doi:10.1016/j.ssci.2008.01.002.

Sachin Kumar Singh and Manoj M. Gadewar, in South African Journal of Botany, 2023

Singh, S.P. D and . Sudhakar Rao V. in Postharvest Biology and Technology of Tropical and Subtropical Fruits: Mangosteen to White Sapote, 2011

Strano M.; Colosimo B.M. (2006). "Logistic regression analysis for experimental determination of forming limit diagrams". International Journal of Machine Tools and Manufacture. **46** (6): 673–682. doi:10.1016/j.ijmachtools.2005.07.005.

Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. **316** (5): 533–4. doi:10.1001/jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067.

Truett, J; Cornfield, J; Kannel, W (1967). "A multivariate analysis of the risk of coronary heart disease in Framingham". Journal of Chronic Diseases. **20** (7): 511–24. doi:10.1016/0021-9681(67)90082-3. PMID 6028270.

Walker, SH; Duncan, DB (1967). "Estimation of the probability of an event as a function of several independent variables". Biometrika. **54** (1/2): 167–178. doi:10.2307/2333860. JSTOR 2333860.

EDITOR-IN-CHIEF
Prof. H. U. Nwanjo [Nigeria]