Analysis of *Salmonella sp* bacterial contamination on Vannamei Shrimp using binary logit model approach

P P Oktaviana¹, K Fithriasari¹
¹Department of Statistics, Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

E-mail: paramitha.oktaviana@gmail.com

**Abstract.** Mostly Indonesian citizen consume vannamei shrimp as their food. Vannamei shrimp also is one of Indonesian exports commodities mainstay. Vannamei shrimp in the ponds and markets could be contaminated by *Salmonella sp* bacteria. This bacteria will endanger human health. *Salmonella sp* bacterial contamination on vannamei shrimp could be affected by many factors. This study is intended to identify what factors that supposedly influence the *Salmonella sp* bacterial contamination on vannamei shrimp. The researchers used the testing result of *Salmonella sp* bacterial contamination on vannamei shrimp as response variable. This response variable has two categories: 0 = if testing result indicate that there is no *Salmonella sp* on vannamei shrimp; 1 = if testing result indicate that there is *Salmonella sp* on vannamei shrimp. There are four factors that supposedly influence the *Salmonella sp* bacterial contamination on vannamei shrimp, which are the testing result of *Salmonella sp* bacterial contamination on farmer hand swab; the subdistrict of vannamei shrimp ponds; the fish processing unit supplied by; and the pond are in hectare. This four factors used as predictor variables. The analysis used is Binary Logit Model Approach according to the response variable that has two categories. The analysis result indicates that the factors or predictor variables which is significantly affect the *Salmonella sp* bacterial contamination on vannamei shrimp are the testing result of *Salmonella sp* bacterial contamination on farmer hand swab and the subdistrict of vannamei shrimp ponds.

1. **Introduction**

Indonesia has huge area which have benefits to the natural resourches for people's welfare, for example aquaculture and mariculture. Ponds is one of the largest aquaculture industry in Indonesia which is indicated by the greatest contribution to the total value of aquaculture production, especially for shrimp and prawn. According to Directorate Generale of Aquaculture in Indonesia, there are two types of shrimps that become main export commodities, tiger shrimp (*Penaeus Monodon*) and vannamei shrimp (*Litopenaeus Vannamei*) [4]. Both of them are highly contributed in export commodities, contains of high protein, and widely consumed by Indonesian society. In fact, vannamei shrimp has the highest increase in average production from 2010 to 2014, which is increase to 20.49% and according to Mahbubillah in [7], this shrimp has high demand both for national and international because of the affordable price and nutritional content. However, vannamei shrimp has high possibility to contaminated by the bacteria, such as *Salmonella sp* [5]. This kind of bacteria in vannamei shrimp can affect people's healthh and causing *salmonellosis* disease.
Based on previous description, the researchers are concern about *Salmonella sp* bacterial contamination on vannamei shrimp and using it as response variables in this research. This study is intended to identify what factors that supposedly influence the *Salmonella sp* bacterial contamination on vannamei shrimp using binary logit model according to the response variable that has two categories.

2. Literature Review

2.1 Binary Logit Model

Binary Logit Model is a model that generated from Binary Logistic Regression analysis which is the analysis method used to find the relationship between the response variable Y that is binary or dichotomous and predictor variables X are polychotomous [3], where this variable X can be qualitative or quantitative. The data type on response variable Y is qualitative with two categories; success (Y=1) and failed (Y=0). Binary logistic regression models are follows [1]:

\[
\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}
\]

Equation (1) can be written log odds or called the logit model that have a relationship linear against x:

\[
g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x
\]

Where \( \frac{\pi(x)}{1 - \pi(x)} \) is odds ratio. If there are \( p \) as independent variables, the logit model is:

\[
g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p
\]

\( g(x) \) is a binary logit model in binary logistic regression.

a. Binary Logistic Regression Parameter Estimation

One method to estimate the parameters of binary logistic regression is Maximum Likelihood Estimation (MLE) [6,10]. The parameters of model being estimated from the vector \( \beta^T = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p) \). The value of the vector \( \beta^T \) obtained by maximise function \( L(\beta) \) through differential of the parameters that will be estimated. Function \( L(\beta) \) is log likelihood function:

\[
L(\beta) = \sum_{j=0}^{n} \left[ \sum_{i=1}^{a} y_i x_{ij} \right] \beta_j - \sum_{i=1}^{n} n_i \ln \left[ 1 + \exp \left( \sum_{j=0}^{p} \beta_j x_{ij} \right) \right]
\]

(4)

This function is a combination of \( Y_1, Y_2, \ldots, Y_n \) which independently, with a value of Y observations consist of success \((Y=1)\) and failed \((Y=0)\), with the binomial distribution and \( E(Y) = n_p(x) \); where \( n_1 + n_2 + \ldots + n_i = N \). For the log likelihood equation in equation (4) is generated from each element of the \( \beta \) and equating with zero. So we obtain \( \sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} x_{ij} \pi(x_i) \). The method to estimate variance and covariance from this parameter estimation is obtained from the second derivative of likelihood function. The second derivative as follows:

\[
\hat{\sigma}^2 \frac{\partial L(\beta)}{\partial \beta^2} = -\sum_{i=1}^{n} x_{ij} \pi(x_i)(1 - \pi(x_i))
\]

(5)

To obtain maximum parameter \( \beta \) on logistic regression equation that could not be solved with MLE method then use Newton Raphson Method. This method is derivation of the Maximum Likelihood non linear method in order to be a form of linear parameters. Parameter \( \beta \) is estimated through iteration as below:

\[
\beta^{(r+1)} = \beta^{(r)} - (H^{(r)})^{-1} q^{(r)}
\]

(6)
where \( t = 0, 1, 2, \ldots \), until convergence:

\[
q^{(t)} = \begin{bmatrix}
\frac{\partial L(\beta)}{\partial \beta_0} & \frac{\partial L(\beta)}{\partial \beta_1} & \cdots & \frac{\partial L(\beta)}{\partial \beta_p}
\end{bmatrix}
\]

\( \mathbf{H} = \begin{bmatrix}
h_{11} & h_{12} & \cdots & h_{1p} \\
h_{21} & h_{22} & \cdots & h_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
h_{p1} & h_{p2} & \cdots & h_{pp}
\end{bmatrix}
\]

Elements of matrix \( \mathbf{H} \) is

\[
H_{ij} = \frac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j}
\]

Algorithm of Newton Raphson iteration process are:

1. Enter the initial value of \( \beta^{(0)} \) to equation:

\[
\beta^{(t+1)} = \beta^{(t)} + \{X^T \text{Diag}(\pi(x_i)^{(t)}(1- \pi(x_i)^{(t)}))X\}^{-1}X^T(y-m^{(t)})
\]

2. Looking for Hessian Matrix \( \mathbf{H}^{(0)} \) and vector \( q^{(0)} \)

3. For \( t > 0 \) then the iteration process followed by applying the equation:

\[
\pi(x_i)^{(t)} = \frac{\exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)}
\]

\[
\beta^{(t+1)} = \beta^{(t)} + \{X^T \text{Diag}(\pi(x_i)^{(t)}(1- \pi(x_i)^{(t)}))X\}^{-1}X^T(y-m^{(t)})
\]

4. Iteration will be stopped if \( \pi(x_i)^{(t)} \) and \( \beta^{(t)} \) have convergent when \( |\beta_j^{(t+1)} - \hat{\beta}_j| < c |\beta_j^{(t)} - \hat{\beta}_j|^2 \), for \( c > 0 \).

**b. Binary Logistic Regression Parameter Test**

There are two kinds of parameters testing in binary logistic regression [6]:

1. **Simultaneous Test**
   
   This test is used to check the significance of the coefficient \( \beta \) simultaneously against the dependent variables. The hypothesis are:
   
   - \( H_0 : \beta_1 = \beta_2 = \ldots = \beta_j = 0 \) (no effect coefficient significantly to model)
   - \( H_1 : \text{at least there is one } \beta_j \neq 0 \), with \( j = 1, 2, \ldots, p \) (coefficient is significantly influence to model)

   The statistical test is \( G^2 \) or **Likelihood Ratio Test**:

   \[
   G^2 = -2 \ln \left[ \frac{n_0}{\prod_{i=1}^{n} \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{1-y_i}} \right]
   \]

   where \( n_0 = \sum_{i=1}^{n} (1-y_i) \) and \( n_1 = \sum_{i=1}^{n} y_i \) and \( n = n_0 + n_1 \)

   Reject the null hypothesis if \( G^2 > \chi^2_{(p-1)} \) where \( p \) is the number of independent variables.

2. **Partial Test**
   
   This test is used to check the significance of the coefficient \( \beta \) partially by comparing alleged \( \beta \) and alleged of standard error. The hypothesis are:
   
   - \( H_0 : \hat{\beta}_j = 0 \) (there is no effect of coefficient to model)
   - \( H_1 : \beta_j \neq 0 \), with \( j = 1, 2, \ldots, p \) (coefficient is significantly influence to model)

   The statistical test is \( W \) (Wald):

   \[
   W = \frac{\hat{\beta}_j}{SE(\beta_j)}
   \]
where $\text{SE}(\hat{\beta}_j) = [\text{Var}(\hat{\beta}_j)]^{1/2}$.

Reject the null hypothesis if $|W| > Z_{\alpha/2}$.

c. **Binary Logistic Regression of Goodness of Fit**

**Goodness of Fit** test of binary logistic regression is used to check if there is significant difference between the observed and model predictions. The hypothesis are:

$H_0$ : model is fit (There is no significant difference between the observed and model predictions)

$H_1$ : model does not fit (There is significant difference between the observed and model predictions)

The statistical test:

$$
\hat{C} = \sum_{i=1}^{g} \left( \frac{a_i - n_i \hat{\pi}_i}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \right)
$$

where:

- $a_i$ = observation on group $k$
- $\hat{\pi}_i$ = average of probability estimation
- $g$ = number of group (combination of categories in simultaneous model)
- $n_i$ = number of observations on group $k$

Reject the null hypothesis if $\hat{C} > \chi^2_{(a,\alpha/2)}$ where $p$ is the number of independent variables.

### 2.2 Salmonella sp Bacteria and Vannamei Shrimp

According to D’Aoust in [5], states that *Salmonella sp* is a short bacteria (1-2 μm), a stem which doesn’t form spores. *Salmonella sp* is capable of fermenting glucose that produces acids and gases. This bacteria actively grows in the pH range from 3.6 to 9.5 and is optimal at a near-normal pH. *Salmonella sp* is one of the most dangerous bacteria in fishery products. These bacteria have a serious impact on human’s health [8]. Vannamei shrimp (*Litopenaeus vannamei*) is a type of shrimp from South America that is widely cultivated in Indonesia since the late 90s to replace the tiger shrimp (*Penaeus monodon*) which has been difficult to cultivate. The advantages of cultivate the vannamei shrimp include: high adaptability to low temperature, relatively fast growth rate, wide market absorption, high survival rate [2]. *A salmonella sp bacterium* is a bacterium that often attacks the vannamei shrimp both in the waters and in the distribution.

### 2.3 Studies from Previous Research (State of the Art)

A study on the pollution of *Salmonella sp* bacteria in vannamei shrimp was done to examined the detection of *Salmonella sp* bacteria on fresh white shrimp (*Penaeus Merguiensis*) in the traditional market of Surabaya Municipality, with the result that fresh white shrimp which marketed in traditional markets in five areas of Surabaya polluted by *Salmonella sp* [9]. The highest pollution rate in South Surabaya (12%) and the lowest pollution in West and North Surabaya (4%). There is also a research examined the technique of *Salmonella sp* bacteria's identification on vannamei shrimp (*Litopenaeus Vannamei*) and its processing water in the development and quality testing laboratory of fishery products in South Sumatera Province with the result that *Salmonella sp* bacteria can cause Salmonellosis disease through food, especially foods that have errors in handling. This situation will provide an opportunity for the microorganisms that cause to grow and move to humans at the time of eating them [5].

### 3. Research Variables

Data that used in this research is the testing result of *Salmonella sp* bacterial contamination on vannamei shrimp. The research variables are shown in Table 1. $Y$ is response variable and $X$ is predictor variable.
Table 1. Research Variables

| Variables | Explanation | Categories | Scale |
|-----------|-------------|------------|-------|
| $Y$       | The testing result of *Salmonella* sp bacterial contamination on vannamei shrimp | $0$ : there is no *Salmonella* sp | Nominal |
|           |             | $1$ : there is *Salmonella* sp |       |
| $X_1$     | The testing result of *Salmonella* sp bacterial contamination on farmer hand swab | $0$ : there is no *Salmonella* sp | Nominal |
|           |             | $1$ : there is *Salmonella* sp |       |
| $X_2$     | The subdistrict of vannamei shrimp ponds | $0$ : Subdistrict A | Nominal |
|           |             | $1$ : Subdistrict B |       |
|           |             | $2$ : Subdistrict C |       |
|           |             | $3$ : Subdistrict D |       |
|           |             | $4$ : Subdistrict E |       |
| $X_3$     | The fish processing unit supplied by | $0$ : PT. X | Nominal |
|           |             | $1$ : PT. Y |       |
|           |             | $2$ : PT. Z |       |
| $X_4$     | The pond are in hectare | - | Ratio |

Data structure of this research is shown in Table 2.

Table 2. Data Structure

| Obs | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|-----|-----|-------|-------|-------|-------|
| 1   | $Y_1$ | $X_{11}$ | $X_{12}$ | $X_{13}$ | $X_{14}$ |
| 2   | $Y_2$ | $X_{21}$ | $X_{22}$ | $X_{23}$ | $X_{24}$ |
| 3   | $Y_3$ | $X_{31}$ | $X_{32}$ | $X_{33}$ | $X_{34}$ |
| 4   | $Y_4$ | $X_{41}$ | $X_{42}$ | $X_{43}$ | $X_{44}$ |
| 5   | $Y_5$ | $X_{51}$ | $X_{52}$ | $X_{53}$ | $X_{54}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $N$ | $Y_n$ | $X_{n1}$ | $X_{n2}$ | $X_{n3}$ | $X_{n4}$ |

4. Result and Discussion

4.1 Descriptive Statistics

The cross tabulation analysis of fish processing unit and testing result of *Salmonella* sp bacterial contamination on vannamei shrimp that is indicated in Table 3; also the cross tabulation analysis of subdistrict of ponds location and testing result of *Salmonella* sp bacterial contamination on vannamei shrimp that is indicated in Table 4, show that there is *Salmonella* sp contamination on vannamei shrimp.

Table 3. The Cross Tabulation Analysis Of Fish Processing Unit And Testing Result Of *Salmonella* Sp Bacterial Contamination On Vannamei Shrimp

| Testing Result | PT. X | PT. Y | PT. Z |
|----------------|-------|-------|-------|
| There is *Salmonella* sp | 2 | 6 | 45 |
| There is no *Salmonella* sp | 3 | 14 | 260 |

Table 4. The Cross Tabulation Analysis Of Subdistrict Of Ponds Location And Testing Result Of *Salmonella* sp Bacterial Contamination On Vannamei Shrimp

| Testing Result | Subdistrict A | Subdistrict B | Subdistrict C | Subdistrict D | Subdistrict E |
|----------------|---------------|---------------|---------------|---------------|---------------|
| There is *Salmonella* sp | 4 | 10 | 8 | 7 | 24 |
| There is no *Salmonella* sp | 46 | 55 | 42 | 53 | 81 |
The results indicate that the problem of *Salmonella sp* contamination on vannamei shrimp is still found and need to cure. That's why this research aim to identify the significant factors that influence *Salmonella sp* contamination on vannamei shrimp using binary logistic model approach.

4.2 Binary Logistic Regression

The first step of binary logistic regression is testing the parameters simultaneously and partially. Simultaneous test indicates to reject $H_0$, p-value = 0.013 (with $H_0$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ and $H_1$: at least one $\beta_j \neq 0$, $j = 1, 2, 3, 4$). It concludes that at least one of predictor variables affects response variable. Partial test concludes to reject $H_0$ for parameters of testing result of *Salmonella sp* bacterial contamination on farmer hand swab and subdistrict of vannamei shrimp ponds. It indicates that there are two significant factors or predictor variables, which are testing result of *Salmonella sp* bacterial contamination on farmer hand swab and subdistrict of vannamei shrimp ponds.

| Table 5. The Results of Binary Logistic Regression |
|-------------|-----|-----|
| Variable    | $\hat{\beta}$ | Odds Ratio |
| (Intercept) | -1.412 | 0.244 |
| $X_1 (1)$   | 1.249 | 3.489 |
| $X_2 (1)$   | -0.482 | 0.618 |
| $X_2 (2)$   | -1.263 | 0.283 |
| $X_2 (3)$   | -0.420 | 0.657 |
| $X_2 (4)$   | -0.798 | 0.450 |

Based on binary logistic regression of response variables and those two significant predictor variables (the results is shown in Table 5), binary logit model could be developed as:

$$\hat{g}(x) = -1.412 + 1.249X_1(1) - 0.482X_2(1) - 1.263X_2(2) - 0.420X_2(3) - 0.798X_2(4)$$

The probability of there is *Salmonella sp* contamination on vannamei shrimp based on two significant predictor variables ($X_1(1)$: there is *Salmonella sp* on worker hand swab; $X_2(1)$: subdistrict B, $X_2(1)$: subdistrict C, $X_2(1)$: subdistrict D; $X_2(1)$: subdistrict E) is 0.0005. The calculation is shown as the following:

$$\hat{g}(x) = -1.412 + 1.249(1) - 0.482(1) - 1.263(2) - 0.420(3) - 0.798(4) = -7.623$$

$$\hat{h}(x) = \frac{\exp(-7.623)}{1 + \exp(-7.623)} = 0.0005$$

From the information, the probability of there is no *Salmonella sp* contamination on vannamei shrimp based on two significant predictor variables ($X_1(1)$: there is *Salmonella sp* on farmer hand swab; $X_2(1)$: subdistrict B, $X_2(1)$: subdistrict C, $X_2(1)$: subdistrict D; $X_2(1)$: subdistrict E) also can be calculated as $1 - 0.0005 = 0.9995$. The contribution of each predictor variable is explained by odds ratio. For example, based on odds ratio of variable $X_1(1)$ (that is predictor variable, category 1 : there is *Salmonella sp* on farmer hand swab) is 3.489. It concludes that the presence of *Salmonella sp* on farmer hand swab tend to cause the absence of *Salmonella sp* bacteria in vannamei shrimp 3.468 smaller than the absence of *Salmonella sp* bacteria on worker hand swab.

5. Conclusion

The result indicates that the predictor variables which is significantly affect the *Salmonella sp* bacterial contamination on Vannamei shrimp are testing result of *Salmonella sp* bacterial contamination on farmer hand swab and subdistrict of vannamei shrimp ponds. The binary logit model could be developed as $\hat{g}(x) = -1.412 + 1.249X_1(1) - 0.482X_2(1) - 1.263X_2(2) - 0.420X_2(3) - 0.798X_2(4)$.
Acknowledgments
We are grateful to LPPM ITS who gives the chance to finish this research in part of the research for beginner (Penelitian Pemula) in Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

References
[1] Agresti A 2002 Categorical Data Analysis (New Jersey: John Wiley & Sons, Inc.)
[2] http://www.agrotaninusantara.com /2014/11/apa-itu-udang-vannamei.html (Accessed on February 20, 2017, at 07.45 pm)
[3] Chen C, Zhang G, Tarefder R, Ma J, Wei H and Guan H 2015 A Multinomial Logit Model-Bayesian Network Hybrid Approach for Driver Injury Severity Analyses in Rear-End Crashes Elsevier : Accident Analysis and Prevention 80 76–88
[4] http://www.djpb.kkp.go.id/arsip/c/246/Udang-Vannamei-dan-Udang-Windu-Masih-Andalan-Ekspor-Indonesia/?category_id=13 (Accessed on February 10, 2017, at 10.16 am)
[5] Hidayat H 2013 Teknik Identifikasi Bakteri Salmonella sp pada Udang Vannamei (Liptopenaeus Vannamei) dan Air Pengolahannya di Laboratorium Pembinaan dan Pengujian Mutu Hasil Perikanan Provinsi Sumatera Selatan (Universitas Sriwijaya: Ilmu Kelautan)
[6] Hosmer D W and Lemeshow S 2000 Applied Logistic Regression, Second Edition (Washington : John Wiley & Sons, Inc.)
[7] Kharisma A and Manan A 2012 Kelimpahan Bakteri Vibrio sp pada Air Pembesaran Udang Vannamei (Litopenaeus Vannamei) sebagai Deteksi Dini Serangan Penyakit Vibriosis Media Jurnal Ilmiah Perikanan & Kelautan 4(2) 20-27
[8] http://mediapenyuluhanperkananpati.blogspot.co.id/2016/09/bahaya-samonella.html (Accessed on February 19, 2017, at 11.51 pm)
[9] Narumi H E, Zuhriansyah and Mustofa I 2009 Deteksi Pencemaran Bakteri Salmonella sp pada Udang Putih (Penaeus Merguiensis) Segar di Pasar Tradisional Kotamadya Surabaya Jurnal Ilmiah Perikanan & Kelautan 1(1) 57-63
[10] Nottorf F 2014 Modeling The Clickstream Across Multiple Online Advertising Channels Using A Binary Logit with Bayesian Mixture of Normals Elsevier : Electronic Commerce Research and Applications 13 45–55