2321-8045 Science Park Research Journal Vol-2, Issue-13, 9th Oct 2014
Impact Factor : 1.6200[UIF-2013]

Original Article

Tests For Binomial And Poission Counts And Rates

M.Ramesh[1] , G.Y.Mythili[2], G.Mokesh Rayalu[3],
HariMallikarjuna Reddy[4], M.Bhupati Naidu[5] and P.Balasiddamuni[6]

ABSTRACT

Biostatistics is a field of Statistical Science in which various statistical techniques may be used for the measurement of biological relationship.In the study of Dose-Response relationships,One may find Binomial and Poission counts and Rates.

In this Research article ,Some tests for equality between Bionomial and Poission counts Rates have been proposed in a simple manner.

Keywords:
Binomial And Poission Counts , Rates , Statistical Science .

M.Ramesh[1] , G.Y.Mythili[2], G.Mokesh Rayalu[3], HariMallikarjuna Reddy[4], M.Bhupati Naidu[5] and P.Balasiddamuni[6]

From
1 Data Scientist, Tech Mahindra , Hyderabad, India
2 Assistant Professor,ACS Medical college and Hospital, Tamilnadu, India
3 Assistant Professor, School of Advanced Sciences,
Statistics and Operational Research Division,
VIT University, Vellore ,Tamilnadu,India
4 Academic Consultant, Department of Statistics, S.V.University,
Tirupati, Andhra pradesh, India
5 Professor, DDE,S.V. University, Tirupati, Andhra pradesh, India
6 Professor, Department of Statistics, S.V.University, Tirupati,
Andhra pradesh, India

The Article Is Published On October 2014 Issue & Available At www.scienceparks.in

DOI:10.9780/23218045/1202013/49

I.INTRODUCTION

Biostatistics is regarded as one of the main branches of statistical science in which various mathematical and statistical methods have been used in Biological Sciences in the widest sense; in Biology, Medicine, Psychology, Agriculture, Forestry, Ecology, Epidemiology and others.

In the comparison of counts or proportions across different populations, it is often important to consider the intrinsic ordering of the populations with respect to some particular characteristic. For instance, one may be interested in assessing whether the proportion of women reporting Insomnia increases with age group or whether the number of accidents is increasing over calendar periods. This type of comparison can be accomplished through the application 'Trend Test'.

II. TREND TESTS FOR BIOSTATISTICS:

In the comparison of counts or proportions across different populations, it is often important to consider the intrinsic ordering of the populations with respect to some particular characteristic. For instance, one may be interested in assessing whether the proportion of women reporting Insomnia increases with age group or whether the number of accidents is increasing over calendar periods. This type of comparison can be accomplished through the application 'Trend Test'. Trend Test arise generally within a wide variety Biostatistical applications, such as

Bioassys, epidemiologic studies and evaluations of environmental exposures etc. in which a Dose-Response relationship may be considered. The characteristic of the population may be measured on a continuous scale, such as an assigned treatment level, or on an ordinal scale (ordered categorical data), such as age group or initial severity of a health condition.

Consider Y_i be a random variable representing the count of interest for the i^{th} population; X_i be quantitative (continuous of ordinal) covariate for the i^{th} population; and w_i be a known design variable for the i^{th} population (often relates to the sample or population size)

Now, $R_i = \frac{Y_i}{W_i}$ represents a rate of a certain event

The form of data for a Trend Test may be given by

Population i	Population covariate X_i	Weight W_i	Observed count Y_i	Rate $R_i = \frac{Y_i}{W_i}$	Expected count $E(Y_i)$
1	X_1	W_1	Y_1	R_1	W1 $f(X_1)$
2	X_2	W_2	Y_2	R_2	W2 $f(X_2)$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
K	X_k	W_k	Y_k	R_k	Wk $f(X_k)$

The expected count relates to the covariate through a continuous function f (x_i) may be written as

$$E\ (Y_i) = W_i\ f\ (X_i) \tag{2.1}$$

One may state the thus null hypothesis as, there is no difference in expected counts due to differences in X_i, so that H_0 may be written as

$$H_0\ f(X_i) = f\ (X_j)\ ,\ \forall\ \ i \neq j\ \ = 1,\ 2,\ .\ .\ .\ k$$

And $H_1 : f\ (X_i)\ \neq f\ (X_i),\ \forall\ i \neq j.$

The one sided alternatives may be written as,

$H_{11}: f(X_i) < f(x_j), \ \forall X_i < X_j$, an increasing trend alternative,

$H_{12}: f(X_i) > f(x_j), \ \forall X_i < X_j$, decreasing trend alternative.

Consider k independent random samples drawn from each of the i = 1, 2, . . ,k populations.

The function f(x) may be considered as either linear function of x say $f(x) = \alpha + \beta$, or a monotone (increasing or decreasing) continuous function of

$\alpha + \beta x$ as $f(x) = g(\alpha + \beta x)$ (2.2)

For instance $g(x) = 1 - \exp[-(\alpha + \beta x)]$ (2.3)

Generally, the inverse function $g^{-1}[f(x)]$ is known as the ' Link function' to be modeled as the linear function $(\alpha + \beta x)$. For example, the link functions for the Normal, Logistic and extreme value models are respectively given by probit, logit and complementary log – log link functions.

By choosing an appropriate model, the null hypothesis may be stated as, $H_0 = \beta = 0$ against

$H_{11}: \beta > 0$, an increasing trend or $H_{12}: \beta < 0$, decreasing trend

For the trend test, any of discrete probability distribution may be assumed for the count random variables y_i.

III. TREND TEST FOR BINOMIAL COUNTS AND RATES

Suppose that, $Y_i \sim$ Binomial distribution and $W_i = n_i$, sample size for the i^{th} population.

Also let $f\ (x_i) = p_i = g(\alpha + \beta x_i)$ (3.1)

One may have, $E\,[Y_i] = n_i\ p_i = n_i\ g(\alpha+\beta X_i)$ (3.2)

The H_0 may be stated as

$H_0 : p_1 = p_2 = . . . = p_k$ 	 $H_{11} : p_1 < p_2 < . . . < p_k$

and $H_{12} : p_1 > p_2 > . . . > p_k$.

To test the null hypothesis, first one may obtain the maximum likelihood estimator for β as follows:

Consider the likelihood function for binomial distribution as

$$L\ (\alpha,\ \beta) = \prod_{i=1}^{k} \binom{n_i}{y_i} p_i^{\ y_i}\ (1- p_i)^{n_i - y_i}$$

$$= \prod_{i=1}^{k} \binom{n_i}{y_i} (\alpha + \beta x_i)^{y_i} \left[1- (\alpha + \beta x_i)\right]^{n_i - y_i} \tag{3.3}$$

Here, g(x) is the identity function which is linear. The maximum likelihood (ML) estimators for α and β may be obtained by solving the following score equations:

$$S(\hat{\alpha}\ \hat{\beta}) = \sum_{i=1}^{k} \begin{bmatrix} 1 \\ x_i \end{bmatrix} \left[y_i - n_i\ \hat{p}_i\right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{3.4}$$

Where $\hat{p}_i = g\ (\hat{\alpha} + \hat{\beta}\ x_i)$ (3.5)

The ML estimator for β is solve by

$$\hat{\beta} = \frac{\sum_{i=1}^{k} x_i\ (y_i - n_i\,\tilde{p})}{\sum_{i=1}^{k} n_i\ (x_i - \bar{x})^2} \tag{3.6}$$

Where $\tilde{\rho} = \frac{\Sigma y_i}{\Sigma n_i}$ and $\overline{x} = \frac{\Sigma\, x_i n_i}{\Sigma n_i}$.

Remarks: (i) For the logistic regression model, pi may be writtenes,

$$p_i = \frac{\exp\left[\alpha + \beta x_i\right]}{1 + \exp\left(\alpha + \beta x_i\right)} \qquad (3.7)$$

In this case, the score equations may be solved by using some methods in the numerical analysis such as the Newton – Raphson or Fisher scoring Algorithm or iterative technique, to obtain the ML estimation $\hat{\alpha}$ and $\hat{\beta}$.

(ii) the link function $g^{-1}f$ (x) may be considered as a second degree polynomial ($\alpha + \beta x + \gamma x^2$) and one can obtain the ML estimations $\hat{\alpha}\ \hat{\beta}$ and $\hat{\gamma}$.

The score test statistic to test $H_0 : \beta = 0$ is given by:

$$Z_{\text{binomial}} = S^{|}\,(\alpha_o, \beta_o)\ I^{-1}\,(\alpha_o, \beta_o)\, S\,(\alpha_o, \beta_o) \qquad (3.8)$$

Where, $\beta_0 = 0$ and I^{-1} (α_0, β_0) is the inverse of the information matrix, evaluated the null hypothesis, one may express,

$$I\,(\alpha, \beta) = \begin{bmatrix} I_{\alpha^2} & I_{\alpha_\beta} \\ I_{\alpha\beta} & I_{\beta^2} \end{bmatrix}$$

$$= -\begin{bmatrix} \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2} & \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha\, \partial \beta} \\ \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \beta\ \ \partial \alpha} & \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2} \end{bmatrix}$$

$$= \sum_{i=1}^{k} n_i p_i\,(1 - p_i) \begin{bmatrix} 1 & x_i \\ x_i & x_1^2 \end{bmatrix} \qquad (3.9)$$

$$= \sum_{i=1}^{k} Var(y_i)\ X_i\ X_i^{|} \qquad (3.10)$$

Where $x_i^{|} = [1\ x_i]$.

$$I^{-|}(\alpha, \beta) = (I_{\beta^2} - I_{\alpha\beta}\ I_{\alpha^2}^{-|}\ I_{\alpha\beta})^{-|}$$

$$\text{or } I^{-|}(\alpha, \beta) = p\ (1-p) \sum_{i=1}^{k} n_i\ (x_i - \bar{x})^2 \qquad (3.11)$$

Now, the score test statistic is given by

$$Z^2_{Binomial} = \frac{S^2(\alpha, \beta)}{I^{-|}(\alpha, \beta)} \qquad (3.12)$$

$$= \frac{[\Sigma\ x_i\ (y_i - n_i\ \tilde{p})]^2}{\tilde{p}\ (1-\tilde{p}) \sum_{i=1}^{k} n_i\ (x_i - \bar{x})^2} \qquad (3.13)$$

In the matrix form, the score test statistic is given by

$$Z^2_{Binomial} = X^{|}\ [Y - E]\ [X^{|}\ V X]^{-|} \qquad (3.14)$$

Where, $X = [(x_1 - \bar{x}),\ \ldots (x_k - \bar{x})]^{|}$,

$$Y = [y_1, y_2, \ldots, y_k]^{|}, \quad E = [n,\ \tilde{p}, \ldots, n_k\ \tilde{p}]$$

And V is the diagonal matrix with elements $n_i\ \tilde{p}\ (1-\tilde{p})$ on the diagonal.

Here, $Z^2_{Bionomial}$ follows asymptotically the χ^2 distribution with one degree of freedom.

The trend test may be frequently used in the analysis of animal bioassay with reference to tumor incidence experiments, in which the animals are randomized to various exposure or dose levels of a drug, chemical or other stimulus and the proportion exhibiting the response of interest is observed. A typical form of bioassay data with binomial counts for lung tumors in female mice exposed 1,2, dichloroethane, for the application of trend test is given by:

Dose (mg/kg)xi	**Number exposed $w_i=n_i$**	**Number with tumor y_i**	**Percentage with tumor (y_i/w_i) 100**
x_0	w_0	y_0	p_0
x_1	w_1	y_1	p_1
x_2	w_2	y_2	p_2

IV. TREND TEST FOR POISSON COUNTS AND RATES:

Consider $Y_1, Y_2, \ldots, Y_k$ be independent poisson random variables and x_i be an ordered covariate assume that $E[Y_i] = W_i\ f(x_i)$. (4.1)

Also consider $f(x_i) = \lambda_i$, the mean of poisson variable y_i.

One may consider the weights W_i arising from one of two situations: either,

(i) Y_i may be the number of rare events during an internal of length W_i, where λi is the event rate per unit time

Or (ii) Y_i may be the sum of Wi independent Poisson random variables, i.e.,

$$Y_i = \sum_{j=1}^{W_i} Y_{ij} \tag{4.2}$$

Where $Y_{i1}, Y_{i2}, \ldots, Y_{iwi}$ are identically distributed with mean λ_i.

For instance, incidence of AIDS or cancer cases per calendar year; number of injuries or accidents over a set time period, number of bacteria per unit volume of suspension; or number of tumors observed in W_i animals exposed to dose x_i in an animal bioassay.

One may test for an increasing or decreasing trend in the means $\lambda_i = E[y_i]/w_i$ with increasing levels of x_i.

The relationship between λ_i and x_i may be specified as

$$\lambda_i = g\,(\alpha + \beta x_i) \tag{4.3}$$

Frequently under poison regression, one may specify

$$g(x_i) = e^{xi} \text{ or } \quad \log(\lambda_i) = \alpha + \beta x_i \qquad (4.4)$$

For the more general specification given by (4.3), the likelihood function may be written as

$$L(\alpha, \beta) = c(Y_1, Y_2, \ldots Y_k) \prod_{i=1}^{k} \exp\{-w_i g(\alpha + \beta x_i)\} \{g(\alpha+\beta x_i)\}^{Y_i} \qquad (4.5)$$

Where, c is a constant independent of α and β. The ML estimators $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by solving score equations, $\frac{\partial \log L(\alpha,\beta)}{\partial \alpha} = 0$

and $\frac{\partial \log L(\alpha,\beta)}{\partial \alpha} = 0$ simultaneously.

As in the case of Binomial counts, iterative numerical analysis methods may be used to obtain the ML estimators $\hat{\alpha}$ and $\hat{\beta}$.

The score test statistic for testing $H_0 : \beta = 0$ is given by

$$Z^2_{poisson} = \frac{\left[\sum_{i=1}^{k} x_i (Y_i - W_i \bar{Y})\right]^2}{\bar{Y} \sum_{i=1}^{k} w_i (X_i - \bar{X})^2} \qquad (4.6)$$

Where, $\bar{Y} = \sum_{i=1}^{k} Y_i \Big/ \sum_{i=1}^{k} W_i$.

Z^2 Poisson follows asymptotically χ^2 distribution with one degree of freedom.

A typical form to data with poisson counts for new cases of melanoma and lung, stomach, reported between a time interval for six age groups, along with the person – years of employment in each age group. In these cases, the variance of poisson counts appears to be inflated relative to the mean.

Age group mid point x_i	Number of observed melanoma cases y_i	Person – years of exposure w_i	Observed rate per 100000 person - years $(y_i/w_i)\ 10^5$	Predicted rate per 100000 person - years $(\hat{\lambda}_i \text{ x } 10^2)$
x_1	y_1	w_1	r_1	$(\hat{\lambda}_1 \text{ x } 10^5)$
x_2	y_2	w_2	r_2	$(\hat{\lambda}_2 \text{ x } 10^5)$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_6	y_6	w_6	r_6	$(\hat{\lambda}_6 \text{ x } 10^5)$

Here the predict rate is based on the fitting of the model, given in (4.3)

V. TEST FOR COMPARING THE EQUALITY OF TWO POISSON COUNTS:

Suppose that Y_1 and Y_2 be the two poisson counts taken over time periods W_1 and W_2 respectively. The two average frequencies or rates are given by $R_1 = (y_1 / w_1)$ and $R_2 = (y_2 / w_2)$ .To test for the equal rates, the test statistic is given by

$$\chi^2 = \frac{[R_1 - R_2]^2}{\left[\frac{R_1}{W_1} + \frac{R_2}{W_2}\right]} \sim \chi_1^2 \qquad (5.1)$$

For large number of counts, the normal approximation is given by

$$Z = \frac{R_1 - R_2}{\sqrt{\left[\frac{R_1}{W_1} + \frac{R_2}{W_2}\right]}} \sim N(0,1) \qquad (5.2)$$

Test For Equality Of More Than Two Poisson Counts

(i) Equal Timings for Poisson Counts

Consider Y_i be the i^{th} count and the same times to obtain the counts are all the same

To test the null hypothesis, $H_0 : Y_1 = Y_2 = . .. = Y_k = Y(say)$ the test statistic is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - \bar{y})^2}{\bar{y}} \sim \chi_{k-1}^2 \qquad (5.3)$$

$$\text{Where } \bar{Y} = \sum_{i=1}^{k} y_i \Big/ k \qquad (5.4)$$

(ii) Unequal Timings for Poisson Counts

Suppose that the time to obtain the i^{th} count y_i be w_i, i = 1, 2, . . . , k. define,

$$\bar{R} = \sum_{i=1}^{k} Y_i \Big/ \sum_{i=1}^{k} wi \qquad (5.5)$$

To test for the equality between the k poisson counts, the test statistic is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{\left[Y_i - W_i \bar{R}\right]^2}{w_i \bar{R}} \sim \chi^2_{k-1} \qquad (5.6)$$

VI. CONCLUSIONS:

In the present study, an attempt has been made by developing some trend tests for biostatistics based on Binomial and Poisson counts and Rates. Also some new tests for equality between the poisson counts have been proposed in the present study.

BIBLIOGRAPHY

1. Altman, D.G, and Bland, J.M. (1983), "Measurement in Medicine: The analysis of medical comparison studies," Statistician 32 :pp 307.
2. Amitage,P.(1955), " Tests for Linear Trends in proportions and frequencies ", Biostatistics,11,PP 375-386.
3. Barnet, R.N. and Youden, W.J, (1970), " A revised scheme for the comparison of quantitative methods," American Journal of Clinical pathology 54 : pp 454-462.
4. Chinn, S. (1990), "The Assessment of Methods of Measurement," Statistics in Medicine, 9: pp 351.
5. Duan, T, Finch, S.J, et.al. (2005), "Using mixture models to characterize disease-related traits". BMC Genet.6 Suppl IS99.
6. Finney, DJ. (1964), "Statistical Methods in Biological Assay," Second Edition, Charles Griffin and company Ltd., London.
7. Hubert, J.J. (1992), "Bioassay", 3rd Ed. Kendall-Hunt, Dubuque.
8. Kanji,G.K.(1999), " 100 Statistical Tests ",Second Edition ,Sage Publications ,London.
9. Kempthorne, O. (1957), "An Introduction to Genetic statistics," IOWA state university press, IOWA, USA.
10. Lee. Y.J. (1980). "Test for trend in count data: multinomial distribution case", Journal of the American Statistical Association 75, 1010 – 1014.

11. Lee. Y.J. (1988), 'Tests for trend in Count Data ",in Encyclopedia of Statistical Science,N.L.Johnson&S.Kotzeds.,New york: Wiley,PP, 328-334