

# Logistic and negative binomial regression models on Factors causing Arthritis

Ms.I. Angel Agnes Mary

Guest Faculty

Department of Statistics

Bharathidasan University, Tiruchirappalli-620023, India.

**Abstract:** Logistic regression, being useful 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. In our study, we find that predicting the outcome of arthritis problem over a long run childhood activities. Negative binomial model, being suitable in situations where the response variable is the 'waiting time' (in discrete units) for an event to happen. In our study, we find that only those who have arthritis problem and wish to build a model that would help in predicting the likely time for onset of arthritis for the risky people, given various explanatory variables.

The objectives of the study are as follows:

- To build a model for predicting the outcome of arthritis problem over a long run.
- To find the expected time for the patients who have the risk of arthritis.

**Keywords:** Arthritis, Logistic Regression, Negative binomial Regression

## I. INTRODUCTION

Arthritis is simply inflammation of the joints. The word rheumatism is more general, and it describes aches and pains in joints, bones and muscles. It is not known whether arthritis generally affects people of all ages, including children. Arthritis is not just part of the ageing process. On the other hand, some kinds of arthritis affects people in particular ages, at the same time, women's are affected by more than men.

**Types:** There are more than 200 different types of arthritis in the world. Some of the arthritis namely, Rheumatoid arthritis, Ankylosing spondylitis, Fibromyalgia, Gout, Polymyalgia rheumatic and Psoriatic arthritis. Many types of arthritis are diagnosed directly, but some types of arthritis are more complex. X-rays and blood tests may be needed to diagnose this. Arthritis usually affects the hips, arms, knees and spine.

**Symptoms:** Significant symptoms of arthritis include pain, fever, tenderness and stiffness, loss of flexibility, tingling sensation, tingling sensation and weight loss. Doctors diagnose arthritis problems using certain methods. They are,

- Medical history
- Physical examination
- Some tests include X-rays, bone scan, CAT scan, MRI, arthroscopy and biopsy.

**Treatments:** There are various treatments available to treat Arthritis. Namely, Exercises, Medications, Meditation, Surgery, Acupuncture, Homeopathy, Chiropractic and osteopathy.

**Some physical activities:** muscle strengthening, walking, exercising in water, swimming or water exercise classes, strength training, yoga, cycling and dancing.

## II. Logistic and negative binomial regression models on Factors causing onset of Arthritis:

### 2.1. Logistic Regression:

Logistic Regression assumes that the outcome variable name is Y, it is categorical variable (dichotomous). Let us consider the binary values of Y as 1 and 0.

Odds in favouring of Y =1 is given by

$$\frac{P(Y = 1)}{1 - P(Y = 1)} = \frac{P(Y = 1)}{P(Y = 0)}$$

$$\text{Odds Ratio} = \frac{\left( \frac{P_i}{1 - P_i} / X_i = x + 1 \right) = e^{\beta_1 + \beta_2(x+1) + u_i}}{\left( \frac{P_i}{1 - P_i} / X_i = x \right) = e^{\beta_1 + \beta_2 x + u_i}} = e^{\beta_2}$$

if odds Ratio < 1,

$$\left( \frac{P_i}{1 - P_i} \middle| X_i = (x + 1) \right) < \left( \frac{P_i}{1 - P_i} \middle| X_i = x \right)$$

if odds Ratio < 1,

$$\left( \frac{P_i}{1 - P_i} \middle| X_i = (x + 1) \right) > \left( \frac{P_i}{1 - P_i} \middle| X_i = x \right)$$

It means that, X is increase and the odds is decreases in favor of Y=1.

if odds Ratio = 1,

$$\left( \frac{P_i}{1 - P_i} \middle| X_i = (x + 1) \right) = \left( \frac{P_i}{1 - P_i} \middle| X_i = x \right)$$

That is, increase in X has no impact on Y.

**2.2. Negative Binomial Regression:**

Suppose there is a sequence of independent Bernoulli trials, each trial having two potential outcomes called “success” and “failure”. In each trial the probability of success is *p* and of failure is (1 - *p*). We are observing this sequence until a predefined number *r* of failures has occurred. Then the random number of successes we have seen, *X*, will have the **negative binomial** (or **Pascal**) distribution:

$$X \sim NB(r : p)$$

**III. Analysis**

**3.1. Logistic Regression:**

Table: 4.1

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	212.630	160.407
SC	215.661	178.589
-2 Log L	210.630	148.407

Table: 4.2

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	D F	Pr > ChiSq
Likelihood Ratio	62.2234	5	<.0001
Score	51.2927	5	<.0001
Wald	35.1174	5	<.0001

Table: 4.3 Results from multivariate logistic regression model containing all explanatory variables

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.9821	2.3474	11.5628	0.0007
Age		1	0.0579	0.0262	4.8905	0.0270
Bmi		1	0.2083	0.1067	3.8126	0.0509
physical_grp		1	-0.5208	0.2852	3.3345	0.0678
food_young_grp		1	-0.5793	0.3162	3.3561	0.0670
sport_grp		1	-1.3662	0.2696	25.6905	<.0001

Table: 4.4

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	1.060	1.007	1.115
Bmi	1.232	0.999	1.518
physical_grp 1 vs 2	0.353	0.115	1.079
food_young_grp 1 vs 2	0.314	0.091	1.084
sport_grp 1 vs 2	0.065	0.023	0.187

Table: 5

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	84.1	Somers' D	0.685
Percent Discordant	15.7	Gamma	0.686
Percent Tied	0.2	Tau-a	0.341
Pairs	5796	C	0.842

**Findings**

- ◆ The odds favoring arthritis increases by 6% as a person progresses one year in age.
- ◆ An increase of one unit of BMI increases the odds for arthritis by 23%.
- ◆ The odds for arthritis decreases by 65% for people who regularly do cycling in their young age.
- ◆ The odd for arthritis is less by 69% for vegetarians as compared to non-vegetarians.
- ◆ Being active in sports in young age reduces the odds for arthritis by approximately 93%.

**4.2. Negative Binomial Regression:**

Table: 4.6

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	63	93.4686	1.4836
Scaled Deviance	63	93.4686	1.4836
Pearson Chi-Square	63	79.9537	1.2691
Scaled Pearson X2	63	79.9537	1.2691
Log Likelihood		3.4953	

Table: 4.7 Results from Negative Binomial regression model containing all explanatory variables

Analysis Of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq	
Intercept	1	2.3152	0.8097	0.7281 3.9022	8.17	0.0042	
gender_grp	1	-0.1980	0.1296	-0.4520 0.0560	2.33	0.1266	
family_history_grp	1	-0.2112	0.1433	-0.4921 0.0697	2.17	0.1405	
physical_grp	1	-0.6200	0.1766	-0.9660 -0.2739	12.33	0.0004	
food_young_grp	1	-0.3774	0.1872	-0.7443 -0.0105	4.06	0.0438	
sport_grp	1	0.4748	0.3051	-0.1233 1.0728	2.42	0.1197	
Dispersion	1	0.1128	0.0483	-0.1667 -0.0182			

Note: The negative binomial dispersion parameter was estimated by maximum likelihood.

Table: 4.8

R Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
gender_grp	1	2.28	0.1310
family_history_grp	1	2.16	0.1420
physical_grp	1	10.35	0.0013
food_young_grp	1	3.26	0.0711
sport_grp	1	3.13	0.0771

**Findings**

- ◆ The expected log-waiting time for onset of arthritis is **0.198** less for females compared to males. That is, on the average,  $e^{-0.198} = 0.82$  is the reduction in the expected waiting time for females compared to males. Converting this time unit into the calendar-year scale, we observe that females tend to develop arthritis about 4 years earlier than a male who is similar in other characteristics.
- ◆ The average expected waiting time  $e^{-0.2112} = 0.81$  where the expected log waiting time for onset of arthritis is **0.2112** high for who are having arthritis in their family.
- ◆ The expected log waiting time for arthritis is **-0.6200** less for the people who are involving physical activities. That is on the average,  $e^{-0.6200} = 0.54$  is the reduction in the expected waiting time for the people who are involving the physical activities.
- ◆ The average expected  $e^{-0.3374} = 0.685$  is the reduction in the expected waiting time for non-vegetarian compared to vegetarian. That is the expected log waiting time for arthritis is **0.3774** less for non-vegetarian compared to vegetarian.
- ◆ The expected log waiting time for onset of arthritis **0.4748** less for the people who being active in sports in young age. That is on the average,  $e^{0.4748} = 1.61$  is reduction in the expected waiting time for the people who being active in sports in young age

## Conclusion

- ❖ Thus, Logistic Regression Analysis is found that Ageing, Increased BMI, Non-vegetarian Food are factors that induced conditions favorable for arthritis. An activity life in the form of doing physical exercise (especially cycling) and involving in sports activities in young age are factors that mitigate the possibilities.
- ❖ Thus, Negative Binomial Regression Analysis is found that, even among people who have risk for arthritis women, people with a family history of arthritis and those who eat non-vegetarian food are likely to suffer earlier onset of arthritis compared to others. Also, physical activeness in the form of cycling and other sports activities bring about a delay in the onset of arthritis.

## Result:

To identify this conclusion,

Risk Factors	Mitigating Factors
Female	Male
Non-Vegetarian	Vegetarian
Sports Activity Absent	Sports Activity Present
Family History of Arthritis	No Family History of Arthritis
Physical Activity Cycling Absent	Physical Activity Cycling Present

## REFERENCES

- [1] American Geriatrics Society Panel on Exercise and Osteoarthritis Work group recommendations: 2002 Exercise and Physical Activity Conference, St. Louis, Missouri. Session V: evidence of benefit of exercise and physical activity in arthritis. *Arthritis Rheum.* 2003;49:453–454.
- [2] American Geriatrics Society Panel on Exercise and Osteoarthritis Exercise prescription for older adults with osteoarthritis pain: consensus practice recommendations. A supplement to the AGS Clinical Practice Guidelines on the management of chronic pain in older adults. [Erratum appears in *J Am Geriatr Soc* 2001 Oct; 49: 1400] *J Am Geriatr Soc.* 2001;49:808–823.
- [3] Rausch Osthoff, AK, Niedermann, K, Braun, J, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. *Ann Rheum Dis* 2018; 77(9): 1251–1260.
- [4] Lee, J, Dunlop, D, Ehrlich-Jones, L, et al. Public health impact of risk factors for physical inactivity in adults with rheumatoid arthritis. *Arthritis Care Res* 2012; 64(4): 488–493.