PLISKA
STUDIA MATHEMATICA BULGARICA
ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

For further information on
Pliska Studia Mathematica Bulgarica
visit the website of the journal http://www.math.bas.bg/~pliska/
or contact: Editorial Office
Pliska Studia Mathematica Bulgarica
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: pliska@math.bas.bg

Pliska Stud. Math. Bulgar. **20** (2011), 39–44

PLISKA
STUDIA MATHEMATICA
BULGARICA

ABOUT THE ORAL HEALTH OF BULGARIAN POPULATION OVER 20 YEARS OLD

Nina Daskalova[1], Boyko Bonev

Caries is the most widespread dental disease around the world. Thus, caries prevalence and intensity are repeatedly studied. Epidemiological research has been conducted covering 1636 persons from Bulgaria. A statistical model for the number of teeth damaged from caries, depending on the age and sex of the patients has been considered in this work.

1. Introduction

Any medical condition disturbs to some extent the individual's capacity, comfort and ability to perform daily activities. This applies fully to the diseases of tissues and organs in the oral cavity and maxillofacial area. Caries is the most widespread dental disease around the world. Thus, caries prevalence and intensity are repeatedly studied. Progress of caries is characterized by irreversibility. After the onset of cavitation in the hard tooth tissue process becomes irreversible. Treatment includes stationing, suspending the progress, and recovering the form and function of teeth through a variety of materials.

Intensity (DMF-T) of caries is a key indicator for assessing the condition of dental hard tissues. It shows the average number of teeth affected by caries of one person from the study group. It is the sum of damaged by caries teeth, not

[1]The research was partially supported by appropriated state funds for research allocated to Sofia University (contract 112/2010), Bulgaria.

2000 *Mathematics Subject Classification:* 62P10, 62J12.

Key words: epidemiological research, caries, statistical data analysis, generalized linear models, logistic regression.

treated (DECAY), missing due to caries teeth (MISSING) and treated, obturated, teeth (FILLING).

An epidemiological survey conducted by Yaneva [5] in 1989 showed that DMF-T for the population over 20 years in Bulgaria is 15.9. A survey of dental health of adults by Yolov [6] in 1996, showed that the DMF-T in those over 60 years of age is 24.25. Boteva, Vassileva and Karova [1] in a comparative study of dental status among the urban population registered a significant increase in the incidence of caries and the number of extracted teeth for a twelve year period in male and female individuals from 18 to 55 years.

Epidemiological research has been conducted covering 1636 persons (894 men and 741 women) over age 20 from Bulgaria. The examined individuals have been selected randomly from six regional cities, including Sofia and six smaller municipalities, for representativeness. The survey consist of detailed clinical examination of the teeth and standardized interview in accordance with WHO recommendations for conducting epidemiological studies since 1997. Project participants have been asked to complete a standard questionnaire to determine their socio-demographic and health status. A statistical model for the number of teeth damaged from caries, depending on the age and sex of the patients has been considered in our work.

2. Descriptive Statistics and Preliminary Data Analysis

Preliminary data analysis has been performed including several steps: detection and management of missing data; data transformations; examination of summary statistics. Later, an appropriate statistical model has been looked for considering linear and generalized linear models and particularly logistic regression. The analysis has been conducted in R (see [2, 4, 3]).

A boxplot of DMF-T by age shows some differences between the levels of age. It is seen that the number of caries teeth increases with age. ANOVA test shows these differences are statistically significant with very small p-value ($< 2e-16$) of the F-test. The chi-square test rejects the null hypothesis of independence with p-value $= 1.56e-11$. The boxplot by gender also shows some difference between men and women and the number for women exceeds the number for men. Here the p-values of ANOVA ($< 2e-16$) and chi-square (0.00678) tests are also small.

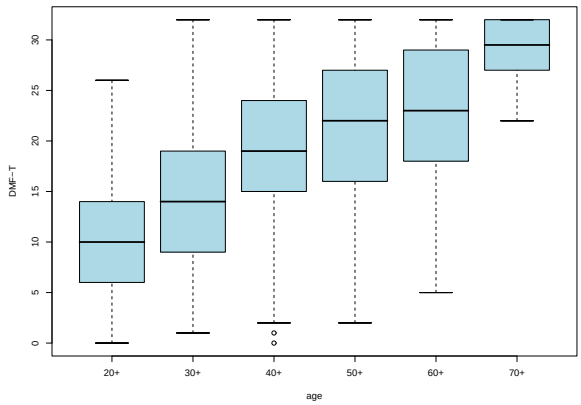

Figure 1: DMF-T by age

3. The Logistic Model

Generalized linear models (GLM) is one of the generalisations of traditional linear models. If we consider a single response variable y and some candidate predictor variables x_1, x_2, ...,x_p, the distribution of y can only depend on the predictors through a single linear function: $\eta = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p$ The distribution belongs to the GLM family of distributions, where there may (or may not) be an unknown scale parameter. There are a number of distributions in the GLM family. It is assumed that the linear predictor determines the mean of the response. The mean is assumed to be a (monotone) function of the linear predictor and the inverse of this function is called the link function. One of the link functions corresponding to the binomial distribution is the logistic link. For the binomial distribution the response is taken as the proportion of cases responding. Thus the mean lies between 0 and 1 and the logistic link uses

$$\mu = \frac{\exp(\eta)}{1 + \exp(\eta)}, \quad \eta = \log \frac{\mu}{1 - \mu}$$

A logistic regression model has been estimated using generalized linear models (the `glm` function) in R. The complete model is

`glm(formula = num.nh.bin ~ age * gend, family = "binomial", data = data.teeth),`

where the response `num.nh.bin` is binomial (the number of unhealthy vs. healthy teeth) and predictors are `age` (continuous) and `gender` (factor) variables.

The coefficient of the interaction effect `age:gend` appears insignificant with p-value 0.152, so a second model with parallel slopes has been considered. The summary is given below.

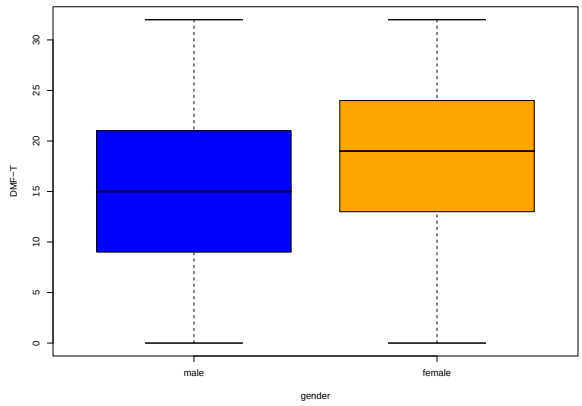

Figure 2: DMF-T by gender

```
    Call:
glm(formula = num.nh.bin ~ age + gend, family = "binomial", data = data.teeth)
Coefficients:
                Estimate Std.Error z-value Pr(>|z|)
(Intercept) -2.0403568 0.0337784 -60.40 <2e-16 ***
age           0.0478705 0.0007756 61.72 <2e-16 ***
gendfemale    0.3756451 0.0184230 20.39 <2e-16 ***
---
Signif.  codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
(Dispersion parameter for binomial family taken to be 1)
Null deviance:  15154 on 1634 degrees of freedom
Residual deviance:  10404 on 1632 degrees of freedom
(1 observation deleted due to missingness)
AIC: 16049
Number of Fisher Scoring iterations:  4
```

This gives coefficients $b_0^m = -2.04$, $b_1^m = 0.05$ for men, and $b_0^f = -2.04 + 0.37 = -1.67$, $b_1^f = 0.05$ for women.

The residual deviance is too large ($P(\chi^2_{1632} > 10404) \approx 0$), so the logistic regression may not provide the best fit. Other models may fit better, though from the scatterplot of the data it is clear that there is great variance in it and most models will suffer from lack of fit.

Nevertheless, it could be seen from the summary that `age` is statistically significant in the logistic regression for the DMF-T and some conclusions can be

made for the slope of the increase of DMF-T. The difference between the levels of the factor variable `gender` could also be inferred. Diagnostics of residuals does not reveal anything to disturb.

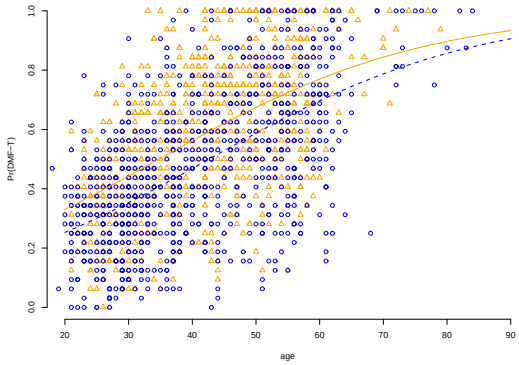

Figure 3: Logistic curves for males (black) and females (grey).

Prop. of DMF-T	Age	SE
p = 0.25:	19.67	0.42
p = 0.50:	42.62	0.26
p = 0.75:	65.57	0.48

Table 1: The proportion of teeth affected by caries for men.

Table 1 shows how the proportion of teeth affected by caries increases with the age according to our logistic model for men. It is seen that at the age of 20, 25% of the teeth has been affected on average, at the age of 43 – the half, and by age of 66 the proportion is 75%. In Table 2 the respective prediction values are given for women.

Prop. of DMF-T	Age	SE
p = 0.25:	12.65	0.52
p = 0.50:	34.95	0.32
p = 0.75:	57.26	0.49

Table 2: The proportion of teeth affected by caries for women.

4. Conclusions

Our analysis shows that the number of teeth affected by caries strongly depends on age and gender. This number increases with age and at any age women have more caries teeth than men (the odds ratio is 1.44). DMF-T of the Bulgarian population over 20 years old has significantly increased since previous research took place.

REFERENCES

[1] E. Boteva, R. Vasileva, E. Karova. Comparative study of dental status among the urban population in Bulgaria. *Problems of Stomatology*, Sofia, 2003, 7–9, (*in Bulgarian*).

[2] J. J. Faraway. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, CRC Press, 2005.

[3] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2010, URL `http://www.R-project.org`.

[4] W. N. Venables, B. D. Ripley. Modern Applied Statistics with S. Fourth Edition. Springer, New York, 2002.

[5] Kr. Yaneva. Stomatological Morbidity and Need of Stomatological Help of Bulgarian Population. Phd. Thesis, Sofia, 1997, p. 60, (in Bulgarian).

[6] Tsv. Yolov. Stomatological Health of Older People in Bulgaria. Phd. Thesis, Sofia, 2000, p. 60, (in Bulgarian).

Nina Daskalova
Sofia University "St.Kliment Ohridski"
Faculty of Mathematics and Informatics
Sofia, Bulgaria
e-mail: `ninad@fmi.uni-sofia.bg`

Dr. Boyko Bonev
Medical University of Sofia
Faculty of Dental Medicine
Sofia, Bulgaria