Statistical Regression Analysis Study on Anorectal Disease Predisposing Factors

Mingliao Niu[1, 2, a], Huanhuan Zhen[1, 2, b], Chunxia Wan[1, 2, c] and Zuowu Xi[1, 2* d]

[1]Henan Province Hospital of TCM, 450000, Henan, P. R. China

[2]The Second Affiliated Hospital of Henan University of Traditional Chinese Medicine, 450000, Henan, P. R. China

[a]hnszyynml@126.com, [b]hnszyyzhh@126.com, [c]wwcx2727@163.com, [d]xizuowu@126.com

*The corresponding author

Keywords: Anorectal disease; Disease-induced factors; Logistic regression analysis

Abstract. Anorectal disease is one of the common diseases of modern society, through research and analysis of its predisposing factors can be targeted to patients with health education, to guide patients to develop good habits, learn self-care, for the prevention of anorectal disease the occurrence of the rehabilitation of patients provide a reference guide. This paper analyzed by Logistic regression showed that constipation, excessive intake of high-protein high-fat foods, sedentary lead to the perineum wet day sitting too long, replace the underwear frequency is too low may lead to the risk factors of anorectal disease, the regular physical exercise, can reduce the incidence of hip bath anorectal disease.

Introduction

Anorectal disease is the particular common disease. Clinically, the most common anorectal disease rational species: hemorrhoids, anal fissure, anal fistula, anal abscess. Anorectal diseases common symptoms are: the anus and abdominal pain, bloody diarrhea, constipation, anal prolapse, anal streaming mucus or bloody fluid bloody stools, difficult defecation, stool shape change ,, anal lump projections, anorectal foreign body sensation, diarrhea , chills, fever, nausea, vomiting, abdominal mass, anemia. Great harm anorectal disease, with the highest incidence of hemorrhoids, a direct impact on people's daily life, work and rest, and even Viagra rib of man's life. Harm mainly as follows [1]: (1) cause anemia: blood in order for the main symptoms of hemorrhoids, often suffered from anemia, if not timely treatment, will appear pale, fatigue, dizziness, weakness, sudden sedentary long after squatting standing, can lead to collapse or fainting. (2) cause skin eczema: Because hemorrhoids prolapse and sphincter relaxation, blood flow outside the anus to stimulate the skin, and cause skin itching and anal eczema. (3) leads to the deterioration of the rectum: Hemorrhoids If not treated in time, the lesions continue to increase, the cell area of the lesion lesions occur, leading to the generation of rectal lesions River. (4) Women's hemorrhoids cause gynecological inflammation: If bleeding hemorrhoids or inflammation often leads to large population of bacteria, causing all kinds of vaginitis, urethritis, cystitis, urinary attachment go far gynecological inflammation.

Logistic Regression Analysis

Logistic Regression. Logistic regression actually belongs discriminant analysis, because it has poor efficiency without discrimination common. It broader range of applications such as: for epidemiological risk factors analysis; laboratory dose of drug - response relationship; clinical trials evaluating; factor analysis prognosis of the disease. According to different classification criteria, divided into different types, as shown in Fig. 1.
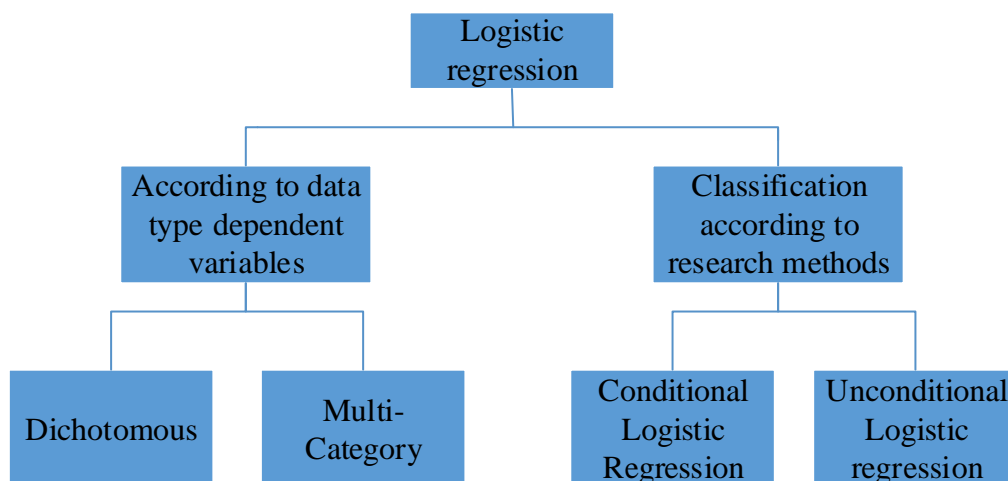
```
                    ┌──────────────┐
                    │   Logistic   │
                    │  regression  │
                    └──────────────┘
              ┌────────────┴────────────┐
   ┌──────────────────┐        ┌──────────────────┐
   │ According to data│        │  Classification  │
   │ type dependent   │        │   according to   │
   │  variables       │        │ research methods │
   └──────────────────┘        └──────────────────┘
     ┌──────┴──────┐             ┌──────┴──────┐
┌───────────┐ ┌──────────┐ ┌───────────┐ ┌───────────┐
│Dichotomous│ │  Multi-  │ │Conditional│ │Unconditional│
│           │ │ Category │ │  Logistic │ │  Logistic │
│           │ │          │ │ Regression│ │ regression│
└───────────┘ └──────────┘ └───────────┘ └───────────┘
```

Figure 1. Logistic Regression Classification

Logistic Regression Model: Let the dependent variable Y is a dichotomous variable whose value is Y = 1 and Y = 0. Effect of Y values of M arguments were $X_1, X_2, \ldots, X_m$. In the M argument (ie, exposure factors) under the effect of the conditional probability of occurrence of a positive result is $P = P(Y = 1 | X_1, X_2, \ldots, X_m)$, the logistic regression model can be expressed as:

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)} \tag{1}$$

Wherein, β_0 is a constant term, $\beta_1, \beta_2, \ldots, \beta_m$ is partial regression coefficient.

Let $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$, the relationship between Z and the logistic curve P is shown in Fig. 2
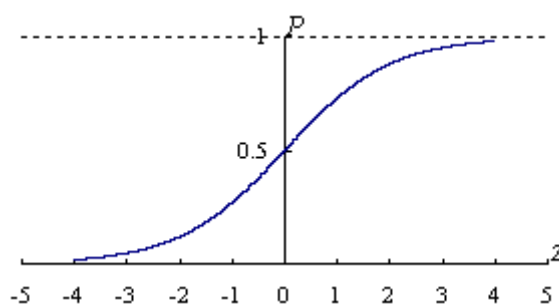

Figure 2. Logistic curve

As can be seen: when z tends to $+\infty$, p incremental value to 1; when z tends $-\infty$, p incremental value to 0; p-value changes between 0 and 1, and with the change in z value to the point (0 0.5) as the center of a symmetrical S-shape.

After as logit transformation, logistic regression model can be expressed in a linear form as follows:

$$\ln\left(\frac{P}{1-P}\right) = \ln\left[\frac{\dfrac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)}}{1 - \dfrac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)}}\right]$$

$$= \ln[\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m)]$$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{2}$$

In the logistic regression, model evaluation index goodness of fit mainly Pearson χ^2, deviation (deviance), Hosmer- Lemeshow (HL) indicator, Akaike Information Criterion (AIC), SC indicators. Pearson χ^2, deviation (deviance) The main argument for the case of small and categorical variables, when the independent variable increases and containing continuous variables, HL index is more appropriate. Pearsonχ^2, deviation (deviance), Hosmer- Lemeshow (HL) index values are subject to χ^2 distribution, χ^2 test was not statistically significant (P> 0.05) represents a good model fit, χ^2 test was statistically significant (P≤0.05) said poor model fit. AIC and SC indicators can also be used to compare the merits of the model, when fitting multiple models, different models and their AIC SC indicator values sorting, AIC and SC values are generally considered less fit better.

Step Fitting Logistic Regression Equation. ① Each quantitative variables and univariate analysis;

② Data discretization for continuous variables in the analysis process is often necessary to become discrete level data. The method can be used are based on experience discrete, or according to quartile, quintile method to determine the level of clustering method can also be used to gather measurement data or class II, into discrete variables.

③ A similar nature to some of the arguments partially multivariate analysis, and explore each variable (level variable, numeric variables) into the model when the appropriate scale, and the arguments are necessary variables transformation;

④ In univariate analysis and related arguments based on the analysis of P≤ α (often taking 0.2, 0.15 or 0.3) of variables, and that the professional is important variables in the multivariate stepwise selection; each fitting a model program models will be given more than one index value for the user to judge the merits of the model and screening variables. You can use two-way filter technology: a variable into the statistics screening score statistics or G or LRS (likelihood ratio statistic), the user determines P value threshold such as: 0.05, 0.1 or 0.2, select the statistic significant and largest variables were entered into the model; b excluding select variables using Z statistics (Wald statistic), the user determines its P value significance level, when the variable is not significant who were excluded from the model [2]. Thus, opt-in and removing cycle repeated until no variables into, and no variable delete it, OK opt or eliminate a significant community of values to be in accordance with the amount of specific issues and variables, generally, when incorporated into the model variable ones, can increase or decrease the value selected into the community exclusion criteria, on the contrary, reduce the value selected into the community, to improve the standard deleted. But different screening criteria will affect the results of the analysis, which results in comparison with others, it should be noted.

⑤ On the basis of multi-factor screening model, based on the need to include consideration of whether the interaction term variable; the interaction between the two variables is an interaction, can be extended to two or more levels of interaction, but in practical applications, the best independent variables (also the requirements of the model itself), do not have to study the interaction of, at most, a small amount of research interactions.

⑥ for professionals consider important but not selected into the regression equation to identify the cause.

Fitting Caution Equation. When equation fitting for Independent Variables Stepwise selection method [method forward (forward), backward method (backward), stepwise regression analysis (stepwise)], the introduction of variable inspection level is less than or equal to exclude variable inspection standards; small sample significance level α set at 0.10 or 0.15, a large sample of the α set at 0.05. Smaller the value the more stringent criteria to select the independent variables; when the regression equation can be selected according to the need to relax or restrict access to the standard equation, or rigid will be the most interested in study variables;

Select influential point record: In theory, the impact of each sample point to the regression model should be the same, is not the case. Some sample point (record) a great influence on the regression model. By the fault or error caused by the point should be deleted without errors and influential point may be related to the relevant arguments and variables can not be easily removed.

multicollinearity diagnosis (SPSS metrics): a tolerance: the more similar to 0, the stronger the collinearity; b characteristic roots: the approximately 0, collinearity stronger; c condition index: the greater the collinearity the stronger; outliers examination: including specific point (outher), high leverage points (high leverage points) and influential point (influential points). Specific point is much larger residuals than the other points of the point; high leverage point is farther away from the other sample points; point is the strong influence of the model have a greater impact point, the model does not contain the point It contains the point regression coefficients obtained will vary greatly. Single specific point or high leverage points will not necessarily affect the estimated regression coefficients, but if both the specific point is the high point of leverage is likely that the impact of a "bad" point of the regression equation. On specific points, the high leverage points, influential point diagnostic indicators are Pearson residuals, Deviance residuals, leverage of statistics H (hat matrix diagnosis), Cook distance, DFBETA, Score test statistics and so on. These five indicators, Pearson residuals, Deviance residuals can be used to check the specific point, if an observed value of residual values> 2, can be considered as a specific point. H lever of statistics can be used to find high leverage point, H is larger samples from the other samples far legend, it can be considered a high leverage point [3]. Cook distance, DFBETA indicators used to measure the specific point or a high degree of leverage points on the regression model. Cook's distance is a standardized synthetic index of both the residual and the degree of leverage, the greater the value, the greater the impact indicates that the corresponding observations. DFBETA index value reflects a change in the regression coefficient Logistic sample is deleted, the larger the change (i.e. DFBETA larger the index value), indicates that the greater the impact of the observed values. If your model check out the specific points, the high leverage points or influential points, first of all should be based on expertise, data collection, analysis and processing, as appropriate, after its causes. Such as from a measurement or recording error correction should be removed or otherwise disposed of have to be cautious, to consider whether to adopt a new model, and not simply deleted even if bin. Because in many cases, abnormal point exactly we detect some clues may be more important factors in advance unclear.

Regression Analysis Method

Object of Study. The object of study for a specific period of time to a hospital in anorectal hospital treatment of patients. Also set up a control group for comparison. The survey of 345 cases of anorectal patients, 157 cases were male, female 188 cases.

Research Methods. The use of questionnaires form, use SPSS 13.0 statistical analysis software, using χ^2 test and multivariate logistic analysis.

Influencing Factors. Can cause anorectal diseases common predisposing factors are smoking, alcohol consumption, constipation, diarrhea, intake of spicy food, high protein intake of high fat foods, fried food intake, daily sitting time, daily standing time , stay up late, sedentary. These predisposing factors were statistically analyzed [4]. Statistical analysis results are as follows:

Table 1 Statistical analysis of predisposing factors

Factors	Case group n%	Control group n%
Smoking	62.5	35.3
Alcohol consumption	61.1	32.5
constipation	61.8	33.4
diarrhea	60.5	34.7
Intake of spicy food	62.1	30.8
Intake of high-protein high-fat foods	61.9	32.8
Intake of fried foods	61.5	30.2
Daily sitting time	62.8	31.5
Standing time daily	62.1	32.5
stay up late	63.9	33.1
Sedentary	64.2	32.8

Using χ^2 test, P <0.001. A description of these factors can cause anorectal disease.

Conclusion

Anorectal disease unique to humans is a common disease, its human health effects is a slow process, due to human anorectal anatomy, the presence of the rectum and anal crypt venous plexus anatomical characteristics, the relative difficulty rectal venous return , prone to cause rectal venous congestion. Therefore, the clinical reason for anorectal disease occurred, all the anatomical structures and physiological characteristics from starting to interpret the role of lifestyle behavior in the anorectal disease, generally only a qualitative description Through this investigation, it is found defecation, eating habits, living habits and so on and the occurrence of anorectal disease.

References

[1] Zhang fang, Ren Boxu, He Xiaobing. Anorectal diseases and behavioral lifestyle case-control study [J]. Modern Preventive Medicine, 2006, (6): 93.

[2] Meler-Ruge wA, Bruder E. Pathology of chronic constipation in pediatric and coloproctology [J] .Pathobiology, 2005, 72 (1-2): 1102.

[3] Chen Ertao. Prevention of anorectal disease [J] .Shanxi Medical College, 2005, 4 (2): 6.

[4] Huang Naijian. China anorectal Science [J]. Shandong Science and Technology Press, 2004.213.