Assignment II

Logistic regression

Silpa Soni Nallacheruvu (19980824-5287) Hernan Aldana (20000526-4999)

2024-11-25

Exercise 2:1

This report analyses the data on periodontitis from a group of adult patients at a large dental clinic. The goal is to understand the influence of the parameter estimates between logistic regression models of the probability for periodontitis in the population as a function of dental floss use and the probability for using dental floss as a function of periodontitis status.

Table 1: Regular Use of Dental Floss and Periodontitis

	Periodontitis	No Periodontitis	Total
Used Dental Floss	22	75	148
Not Used Dental Floss	265	97	413
Total	170	340	510

Question 1: Logistic regression model of the probability for periodontitis

The logistic regression model of the probability for periodontitis in the population as a function of dental floss use is defined as:

$\text{logit}(p_x) = \log\left(\frac{p_x}{1-p_x}\right) = \beta_0 + \beta_1 x$

where:

- $x = 1$ if dental floss is regularly used, $x = 0$ otherwise.
- $p_x = P(\text{periodontitis} \mid x)$, the probability of periodontitis given x.
- β_0: The log-odds of periodontitis when $x = 0$ (no floss use).
- β_1: The change in log-odds of periodontitis when x changes from 0 to 1 (effect of using floss).

Approach:

Fit the logistic regression model to the data and interpret the parameter estimates.

Estimate Parameters:

1. Fit the logistic regression model:

The summary of the logistic regression model coefficients is presented below:

Table 2: Summary of Logistic Regression Model Coefficients

	Term	Estimate	Std. Error	z value	P-value
(Intercept)	(Intercept)	-0.5825176	0.1026175	-5.676593	0.0000000
floss	floss	-0.6439281	0.2632835	-2.445759	0.0144548

2. Expected Estimates of the parameters:

2.1. Intercept (β_0):

- β_0 represents the log-odds of periodontitis when $x = 0$ (no dental floss use): $\beta_0 = \log\left(\frac{p_0}{1-p_0}\right)$ where: $p_0 = \frac{\text{Periodontitis (No Floss)}}{\text{Total (No Floss)}} = \frac{148}{413} \approx 0.3585$
- Hence: $\beta_0 \approx \log\left(\frac{0.3585}{1-0.3585}\right) = \log(0.558) \approx -0.583$

2.2. Slope (β_1):

- β_1 represents the change in log-odds when $x = 1$ (dental floss is used): $\beta_1 = \log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_0}{1-p_0}\right)$ where: $p_1 = \frac{\text{Periodontitis (Floss)}}{\text{Total (Floss)}} = \frac{22}{97} \approx 0.2268$
- Hence: $\beta_1 \approx \log\left(\frac{0.2268}{1-0.2268}\right) - \log\left(\frac{0.3585}{1-0.3585}\right) \approx \log(0.293) - \log(0.558) = -1.229 + 0.583 = -0.646$

The final logistic regression equation: $\text{logit}(p_x) = -0.583 - 0.646x$

Interpret Parameters:

- 1. β_0 = -0.583:
  - The log-odds of periodontitis for individuals who do not use dental floss is approximately -0.583.
  - The corresponding probability of periodontitis is: $p_0 = \frac{e^{-0.583}}{1+e^{-0.583}} \approx 0.358$
- 2. β_1 = -0.646:
  - The log-odds of periodontitis decreases by 0.646 when individuals use dental floss regularly.
  - This corresponds to an odds ratio of: $\text{Odds Ratio} = e^{\beta_1} = e^{-0.646} \approx 0.524$

Individuals who use dental floss regularly have approximately 52.4% lower odds of developing periodontitis compared to those who do not.

Conclusion:

- The logistic regression model provides interpretable estimates of the effect of dental floss use on periodontitis.
- Regular dental floss use is associated with a significant reduction in the odds of periodontitis.
- The logistic regression model is appropriate because the data is binary (periodontitis: yes/no) and the explanatory variable (dental floss use) is categorical.

Question 2: Logistic regression model of the probability for using dental floss

Approach :

R Output :

Observation :

Question 3:

Approach :

R Output :

Observation :

Exercise 2:2

Approach :

Code :

Output :

Observation :