**Statistical Report Writing Sample No.7.**

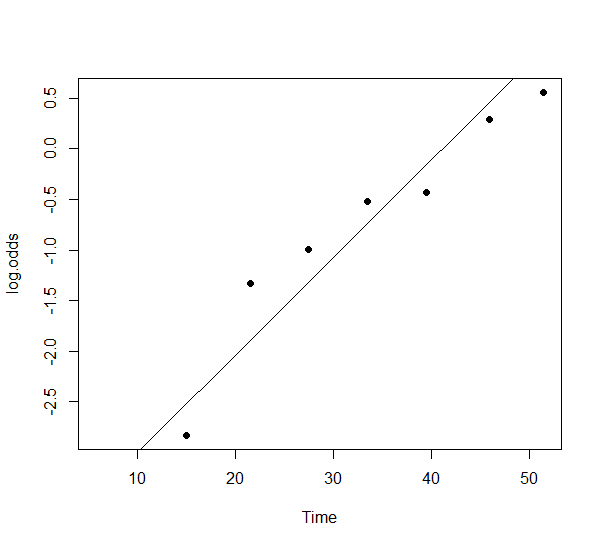
**Introduction.** A study on the degree of pneumoconiosis in coalface workers were conducted in order to examine the relationship between exposure time (in years) and degree of disease. Severity of disease was originally rated into three categories, but here we will use only two (normal and diseased). The following table shows the period of exposure (5.8, 15, 21.5 years, so on) and the logarithmic transformation of the time.

|  |  |  |  |
| --- | --- | --- | --- |
| Time | Log.time | Normal | Diseased |
| 5.8 | 1.757858 | 98 | 0 |
| 15 | 2.70805 | 51 | 3 |
| 21.5 | 3.068053 | 34 | 9 |
| 27.5 | 3.314186 | 35 | 13 |
| 33.5 | 3.511545 | 32 | 19 |
| 39.5 | 3.676301 | 23 | 15 |
| 46 | 3.828641 | 12 | 16 |
| 51.5 | 3.941582 | 4 | 7 |

**Data analysis.**  We can consider the logistic regression with the original time as a covariate, and obtain the following table to summarize the result.

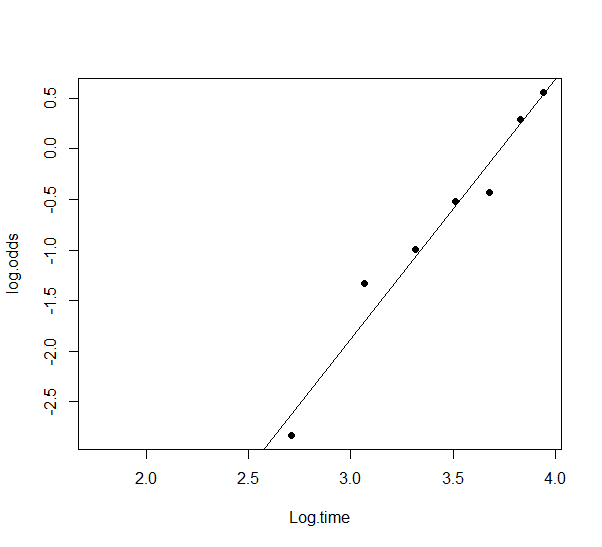
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor | Estimate | L.Bound | U.Bound | p.value |
| (Intercept) | -3.966352 | -4.844063 | -3.195977 | 2.86E-21 |
| Time | 0.09626925 | 0.07316471 | 0.1218069 | 6.90E-15 |

The pair of lower bound and upper bound in the table indicates the 95% confidence interval for the respective parameters. The log odds for the covariate Time is highly significant, and the odds ratio of exp(0.096) ≈ 1.1 indicates that the odds increase by 10% in one year. For example, if the exposure time is doubled from 20 years to 40 years, the odds increase to (1.1)^(20) ≈ 6.7.



The scatter plot for Time against the log odds shows concavity downward, and suggests the use of logarithmic transformation. When we applied the logarithmic transformation of exposure time as a covariate, and obtained the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor | Estimate | L.Bound | U.Bound | p.value |
| (Intercept) | -9.608919 | -12.47122 | -7.215371 | 7.19E-13 |
| Log.time | 2.576021 | 1.879973 | 3.396489 | 2.59E-11 |



The scatter plot for the logarithmic transformation reasonably indicates a linear relationship between the log time and the log odds, and the slope for the logarithmic transformation of exposure time is highly significant. The odds of exp(2.57) ≈ 13.1 indicates that the odds are 13.1 times higher in the unit increment of logarithmic transformation. For example, if the exposure time is doubled, the odds increase to (13.1)^log(2) ≈ 5.9.

**Conclusion.** The logistic regression analysis shows that the exposure time increases the risk of pneumoconiosis. The slope is significant for the original exposure time or its logarithmic transformation, but the scatter plot indicates a better linear relationship when the logarithmic transformation was used. The model can predict the chance of developing the disease at certain exposure time. For example, the coalface worker with 30 years of exposure has the estimated probability of 25% if you use the model with original exposure time, or 30% if you consider the logarithmic transformation. In either of the models about 41 years of exposure time gives 50% chance of having developed the disease.