**Exercises 6 Solutions**

1.

a. H₀ : There is no relationship between years of education and the response to the question,β₁=0.

HA : There is a relationship between years of education and the response to the question,β₁≠0.

b. p represents the proportion of people that agreed with the statement.

c. Yes the model does indicate that there is a significant relationship between the years of education a person has and their response to the statement. The estimate for the coefficient determining the strength of the relationship (β₁=-0.27065) is highly significant (<2×10⁻¹⁶).

d. ln((p/(1-p))) = 2.50334-0.27065x

e. p = 0.3220296

2.

* 1. The distribution of the antibody variable had a strong positive skew, the log transformation corrected this skew.

* 1. Yes the model does indicate that there is a significant relationship between the antibody level and the presence of malaria. The estimate for the coefficient determining the strength of the relationship (β₁ = -0.68235) is highly significant (p=0.000483)

H₀ : There is no association between Age and the probability that an individual will show symptoms of Malaria β2=0.

HA : There is an association between Age and the probability that an individual will show symptoms of Malaria β2≠0.

The residual deviance for Model 2 is 98.968 with 98 degrees of freedom. The residual deviance for Model 1 is 98.017 with 97 degrees of freedom. The difference between the residual deviance for Model 1 and Model 2 is 0.951 with 1 degree of freedom. From the statistical table, the critical value at the 5% level with 1 degree of freedom is 3.841, therefore the calculated value is less than the critical value and Model 1 and Model 2 are not significantly different. We fail to reject the null hypothesis and conclude that Age is not associated with the probability that an individual will show symptoms of Malaria.

where p represents the probability that an individual will show symptoms of malaria, has the value 2.1552 , has the value -0.7122 and represents the antibody level.

* 1. Calculate the probability that an individual with an AB level of 300 will show symptoms of malaria.
  2. The residual deviance for Model 2 is 98.968 with 98 degrees of freedom. The residual deviance for the Null Model is 116.652 with 99 degrees of freedom. The difference between the residual deviance for Model 2 and the Null model is 17.684 with 1 degree of freedom. For the statistical table, the critical value at the 5% level with 1 degree of freedom is 3.841, therefore the calculated value is greater than the critical value and Model 2 is significantly better than the Null Model. We should select Model 2

1. How many principal components are needed if you wish to account for 95% of variation in the data? Please round the output data to 2 decimal places.

Two principal components are required to account for 95% of the variation in the data.

1. Based on the output of the PCA, is it possible to distinguish between the soil samples? Please explain your answer, referencing the output you used to reach your decision.

Yes it is possible to distinguish between the three categories of soil samples. The samples from categories A, B and C are separated on the scores plot shown in Figure 2.

1. Comment on the scores plot shown in Figure 2.

The samples from category A were all positively correlated with PC1 and PC2, with the exception of one sample which had a very weak negative correlation with PC1.

The samples from category B were clustered near the origin.

The samples from category C were spread out on the scores plot, some negatively correlated with both PC1 and PC2 and the remainder negatively correlated with PC1 or PC2. No samples from category C were positively correlated with both PC1 and PC2.

Both PC1 and PC2 are required to distinguish between the samples.

1. Which category of soil samples has comparatively high levels of the trace elements P, Ca, Zn, Cu and Mn?

Category A

1. Comment on the composition of trace elements found in soil samples that were subject to a high level of human activity in the past.

Soil samples from category A were subject to the highest level of human activity in the past and Figures 2 and 3 show that these have comparatively high levels of the trace elements P, Ca, Zn, Cu and Mn.

1. The scree plot shown in Figure 1 has no labels on the axes. What does the *y*-axis measure?

The y axis measures the amount of variation accounted for by each principal component.

1. Based on the scree plot shown in Figure 1, how many principal components would you retain?

Two principal components.

1. Which variables (trace elements) are well explained by the second principal component?

Fe, V, Co, Cr