

SNAIL MORTALITY STUDY

Arden Miller

Executive Summary

For any fixed levels of temperature, relative humidity and exposure time within the ranges used in the study, the predicted odds of mortality for snail species B is 3.67 times the predicted odds of mortality for a snail species A. For both species, there is a dramatic increase in the odds of mortality as exposure time increases from 1 week through 4 weeks. The odds of mortality increases as temperature increases from 10C through 20C, and decreases as relative humidity increases from 60 through 75.8.

1 The Data

This data comes from an experiment that was conducted to investigate the mortality rates of two species of snails under different combinations of temperature and relative humidity. Groups of 20 snails were held for periods of 1, 2, 3 or 4 weeks in carefully controlled conditions of temperature and relative humidity. All possible combinations of three different temperatures (10, 15, and 20 degrees Celsius) and four different levels of relative humidity (60, 65.8, 70.5, and 75.8) were used. At the end of the exposure time the snails were tested to see if they had survived.

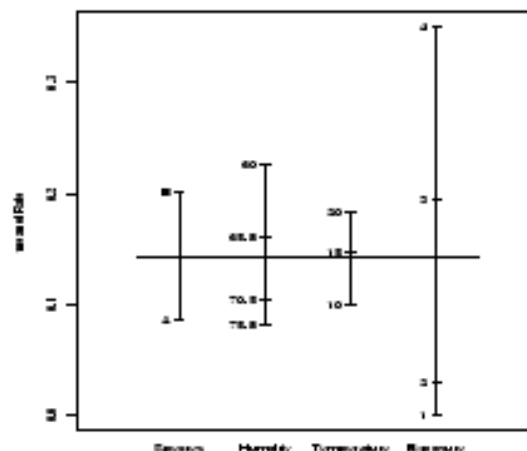


Figure 1: Mean mortality rates

Figure 1 compares the mean mortality rates for different levels of the explanatory variables. This plot indicates that on average the observed mortality rates were higher for species B than for species A. Further it shows that the mortality rates tend to increase as relative humidity decreases, as temperature increases, and as exposure time increases. The biggest differences in mortality rates are observed for the different exposure times.

Figure 2 contains trellis plots of the observed mortality rates for each group of snails versus exposure period. Each panel represents a different combination of temperature and relative humidity. The top set of plots are for species A and the bottom set for species B. These plots indicate that mortality rates are generally higher for species B than for species A. For both species no snails died in the first week for any combination of temperature and relative humidity and the observed mortality rate increases as exposure time increases. For both species the observed mortality rate increases as temperature increases from 10C through 20C and the observed mortality rate decreases as relative humidity increases from 60 through 75.8.

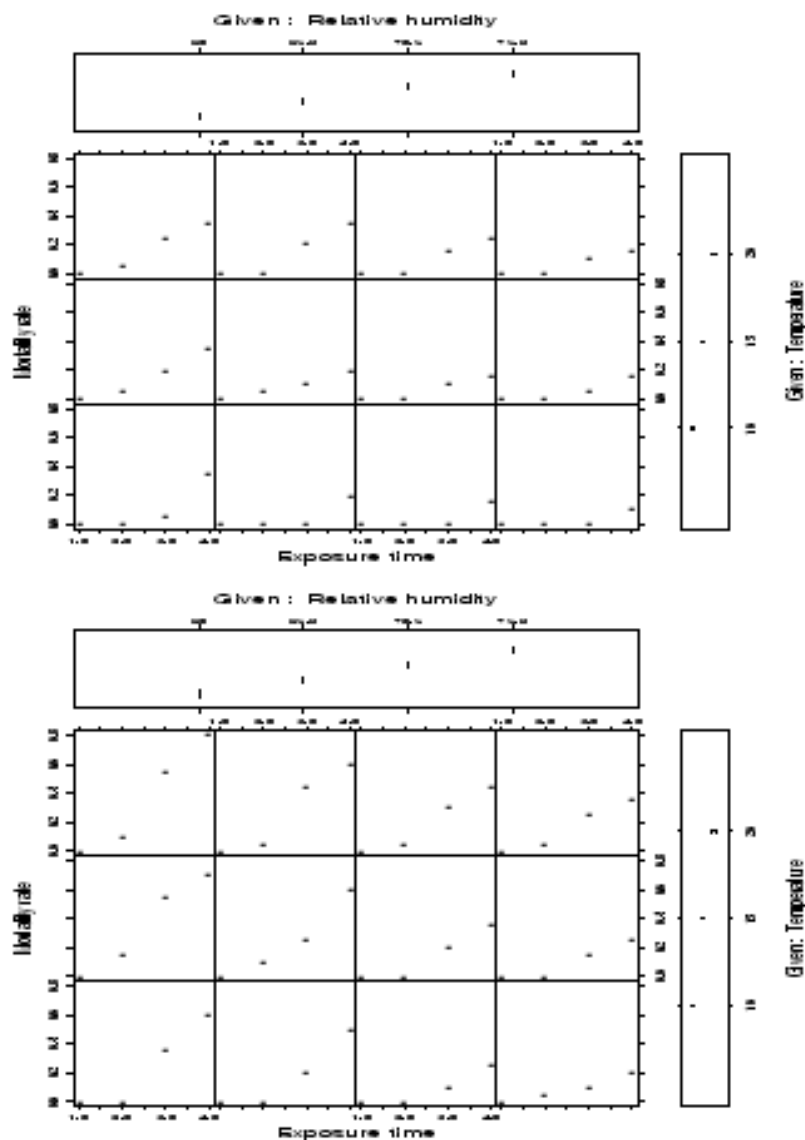


Figure 2: Trellis Plots: top-Species A bottom-Species B

2 A Model for Snail Mortality Rates

A logistic regression model was used to model the probability of mortality for snails (π) using species, exposure time (**Exp**), temperature (**Temp**), and relative humidity (**RH**) as explanatory variables. This logistic regression model relates the log odds of mortality, $\log(\pi/(1 - \pi))$, to the levels of the explanatory variables. The fitted model is

$$\log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = 7.549 + 1.299 \text{ Species} - 0.106 \text{ Humid} + 0.093 \text{ Temp} - 12.473 \text{ Exp}^{-1}$$

For this model **Species** takes the value 0 for species A and the value 1 for species B, **Humidity** represents values of the relative humidity (values should be from 60 to 75.8), **Temp** represents temperature in degrees Celsius (values should be from 10 to 20), and **Exp** represents exposure time in weeks (values should be from 1 to 4). A reciprocal transformation of exposure time was used as this noticeably improved the model fit.

By taking the exponential of both sides of our model, we can get a model for the odds of snail mortality.

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(7.549) \times \exp(1.299 \text{ Species}) \times \exp(-0.106 \text{ Humid}) \times \exp(0.093 \text{ Temp}) \times \exp(-12.473 \text{ Exp}^{-1})$$

The odds of snail mortality is the ratio of the probability a snail dies to the probability it survives. Therefore if the odds of mortality is 1 it indicates a snail is equally likely to die as survive whereas an odds of mortality of 5 indicates that a snail is 5 times as likely to die as survive.

We can use the fitted model to investigate the effect that each of the different explanatory variables have on the odds of snail mortality. Notice that each of the terms on the right hand side of our model only contains one of the explanatory variables. So if we hold all of the explanatory variables constant except one, then all the terms will remain constant except the term involving that explanatory variable. In this way we can isolate the effect of that variable on the odds of snail mortality.

First, consider the effect of **Species** on snail mortality. If temperature, relative humidity, and exposure time are fixed then the difference in the odds of mortality for the two species occurs because the second term in our model is $\exp(0) = 1$ for species A and $\exp(1.299) = 3.67$ for species B. This indicates that the odds of mortality for species B are 3.67 times the odds of mortality for species A. This is true for any combination of temperature, relative humidity, and exposure time.

If species, temperature, and exposure time are fixed then for each unit increase in relative humidity the odds of mortality is predicted to decrease by a factor of $\exp(-0.106) = 0.90$ (it becomes 90% of what it was previously).

If species, relative humidity, and exposure time are fixed then for each increase of 1C in temperature, the odds of mortality is predicted to increase by a factor of $\exp(0.093) = 1.10$ (a 10%

increase).

A bit of care needs to be exercised in interpreting the results for exposure since we have used the reciprocal transformation. Our model predicts that if species, temperature, and relative humidity are fixed then for each unit increase in Exp^{-1} the probability of mortality decreases by a factor of $\exp(-12.473) = 0.0000038$. In order to relate this to changes in exposure time, suppose we will treat an exposure time of 4 weeks as a baseline and then compare the odds of mortality at exposure times of 1, 2 and 3 weeks to the odds of mortality at 4 weeks. Our model predicts the following:

$$\begin{aligned}\text{odds at 1 week} &= 0.00009 \times \text{odds at 4 weeks} \\ \text{odds at 2 weeks} &= 0.04423 \times \text{odds at 4 weeks} \\ \text{odds at 3 weeks} &= 0.35366 \times \text{odds at 4 weeks}\end{aligned}$$

This shows a dramatic increase in the odds of mortality as the exposure time is increased from 1 week through 4 weeks.

Statistical Appendix

For this assignment it was first necessary to identify a suitable logistic regression model that relates the probability of snail mortality to the explanatory variables. The data is from a designed experiment. Three of the explanatory variables (**Exp**, **RH**, and **Temp**) are numeric regressors but only a small number of distinct levels were used for each of these. Therefore they could be treated as factors. Initially, I decided I would treat **Exp**, **RH**, and **Temp** as factors and investigate the need for interactions. Then later, I investigated simplifying the model by using **Exp**, **RH**, and **Temp** as numeric regressors.

First, I created factors for all of the explanatory variables and fitted the full (maximal) model. None of the interactions came up as being significant. A Chi-square test to test whether all interactions can be dropped has a test statistic of $\chi^2_0 = 30.82$ and 86 degrees of freedom. The corresponding P-value is essentially 1 and so gives no evidence that the interactions are needed.

Next I fitted a model that just had linear effects for **Exp**, **RH**, and **Temp** and tried adding the quadratic terms. The quadratic for **Exp** was highly significant but the quadratic terms for **RH** and **Temp** were not at all significant. So only the quadratic term for **Exp** was retained.

Finally I tried transforming **Exp** using the power family of transformations to see if I could eliminate the need for a polynomial model in **Exp**. I found using $\lambda = -1$ gives a residual deviance of 32.7 as compared with 32.9 for using a quadratic function of **Exp**. These are very similar but I decided to use the power transformation model since it has 1 less fitted coefficient and will be somewhat easier to explain.

The S-plus output for my selected model is:

Value	Std. Error	t value
-------	------------	---------

```

(Intercept)  7.54917295  1.02664774  7.353226
Species      1.29113398  0.16177507  7.981044
RH          -0.10568597  0.01378192 -7.668452
Temp         0.09323828  0.01916605  4.864761
I(Exp^-1)  -12.47267856  1.08136383 -11.534211

```

(Dispersion Parameter for Binomial family taken to be 1)

Null Deviance: 539.7207 on 95 degrees of freedom

Residual Deviance: 32.69197 on 91 degrees of freedom

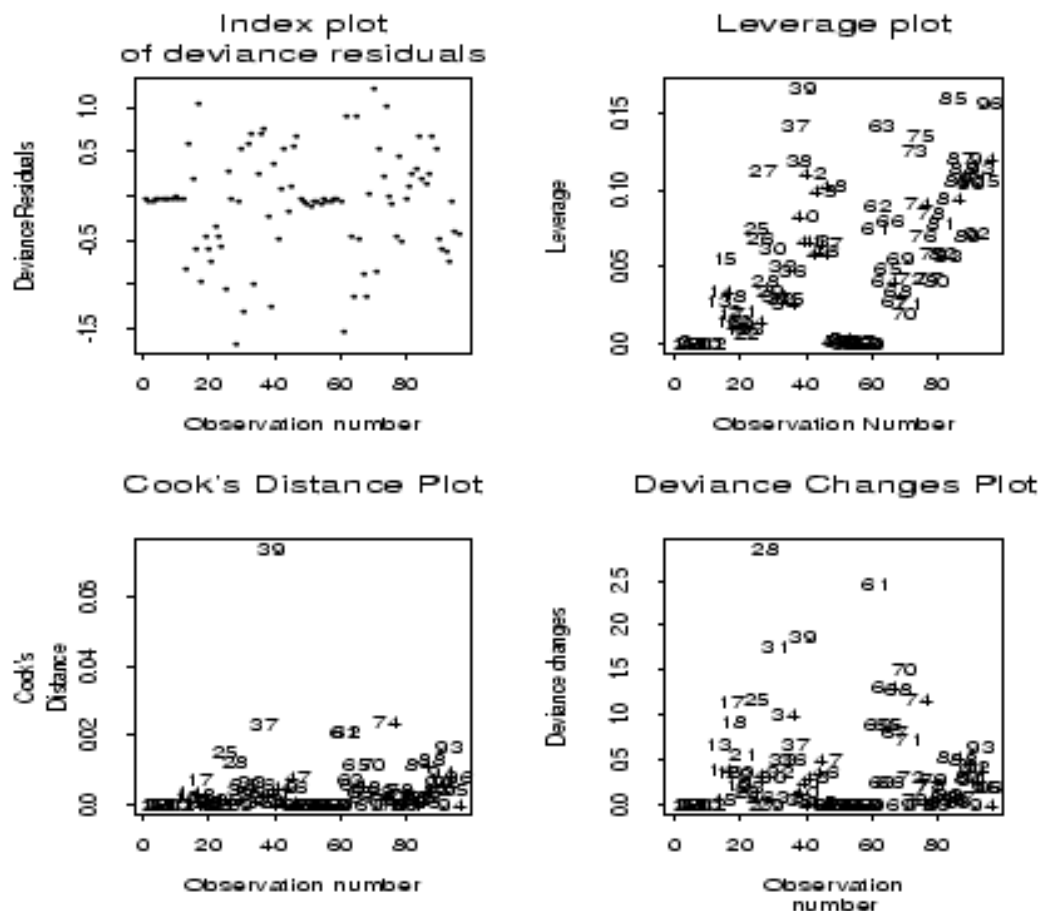


Figure 3: Diagnostic plots for final model

The only point that shows up as being influential in Figure 3 is observation 39 due to a somewhat large Cook's Distance. Observation 39 corresponds to Species A, 4 weeks exposure time, a temperature of 60C and relative humidity of 60. If we identify this point on the trellis plots in Figure 2, it does not stand out as being unusual. The values of DFBETAS and DFFITS for this point are:

```

(Intercept)  Species      RH      Temp  I(Exp^-1)  dffits
-0.5663353  0.4719847  0.5798798 -0.5521505  0.3948727 -1.144255

```

All the DFBETAS values are in the order of .5 so if this observation is dropped the fitted coefficients would each change by about 0.5 standard errors. This is not a big effect and would not affect the general conclusions obtained using the fitted model. The DFFITS value indicates that $\text{logit}(\hat{\pi})$ would increase by 1.1 standard errors if this point were deleted.

Since the impact of deleting observation 39 is minimal and there is no evidence that it is unusual, I have only reported the results for the full data set.

475.330 Assignment 4: Marking Guide

This assignment asks the students to analyse data from an experiment that investigated the mortality rates of 2 species of snails under different conditions.

Report	14 marks
Statistical Appendix	6 marks
Total	20 marks

Report for biologist interested in snail mortality (14 Marks)

This part of the assignment should describe their findings in terms that a non-statistician can understand. They should identify a suitable model for the probability of mortality for snails and clearly explain what this model indicates.

- Presentation - 5 marks for a generally well laid out, coherent report. The reader should not have to search for the important parts among a lot of details. Look for good use of graphs and clear explanations of the way the numerical regressors affect risk and a clear discussion of differences between regions. At some point in their report (i.e. an Executive Summary) they should summarise their main findings in a short paragraph. Give: (1) 5 marks for a clear, precise report, that is easy to follow (2) 3 marks if the report is difficult to follow or contains a lot of unnecessary detail, or would be difficult for a non-statistician to understand, (3) 1 mark if it would be very difficult for anyone to understand.
- Content - 9 marks for a report that accomplishes the following
 1. Identifies a suitable logistic regression model for snail mortality.
 - The model should not contain any interactions.
 - The variables **Exp**, **RH** and **Temp** should be treated as numeric regressors in their final model as this makes the model more useful than treating them as factors.
 - The model is improved if **Exp** is transformed. Any power transformation from $\lambda = 0$ (log) to $\lambda = -2$ is reasonable. Alternatively, they may have used a quadratic model for **Exp** - this is fine but a cubic term in **Exp** is definitely not needed. No transformations are needed for **RH** or **Temp**.
 2. Explains how the model relates the explanatory variables to probability of mortality in snails. It is not enough to simply write down a model such as

$$\text{logit}(\hat{\pi}) = 7.549 + 1.299 \text{ Species} - 0.106 \text{ Humid} + \dots$$

They would need to explain what $\text{logit}(\hat{\pi})$ means and explain how to convert values of $\text{logit}(\hat{\pi})$ to values of $\hat{\pi}$ (or to odds provided they explain what odds means).

3. They need to discuss the impact that each of the explanatory variables (**Species**, **Exp**, **RH** and **Temp**) has on snail mortality. Pay particular attention to the way they explain the impact of **Exp** since it will have been transformed (or be a quadratic function) - make sure their explanation makes sense.
4. Look for the use of a few well chosen plots in their report.

Statistical Appendix (6 Marks)

This appendix should outline the reasons that they came to the conclusions they presented in the first part of the analysis. They are not required to give a detailed account of everything they did but they should present a coherent account of their analysis. They should do the following:

1. Identify a suitable logistic regression model for this data.
 - They may have treated **Exp**, **RH** and **Temp** as factors during the initial stages of model selection (when considering the need for interactions) but should treat them as numerical variables in their final model.
 - At some point they should indicate that they investigated adding interactions (at least 2-way interactions) but found this unnecessary.
 - They should also have investigated transforming (or using polynomial models for) the 3 numerical regressors (**Exp**, **RH** and **Temp**) either using plots (partial plots and constructed variable plots) or by trial and error. They should find that only **Exp** needs to be transformed.
2. Do diagnostic plots for their chosen model
3. Identify influential points and assess what effect deleting these points has on the fitted model.

Note:

- Include short comments indicating why a student has lost marks.
- Include a break up of marks into presentation, content, and statistical appendix.
- There were problems with printing in the advanced lab. As a result, some reports may have the plots on separate pages and not in the text of the report. This is fine.