

02424 Week 7

Today you should be working on an exercise, which considers the possibility of having a number of explanatory variables or covariates in a generalized linear model.

Survival of snails

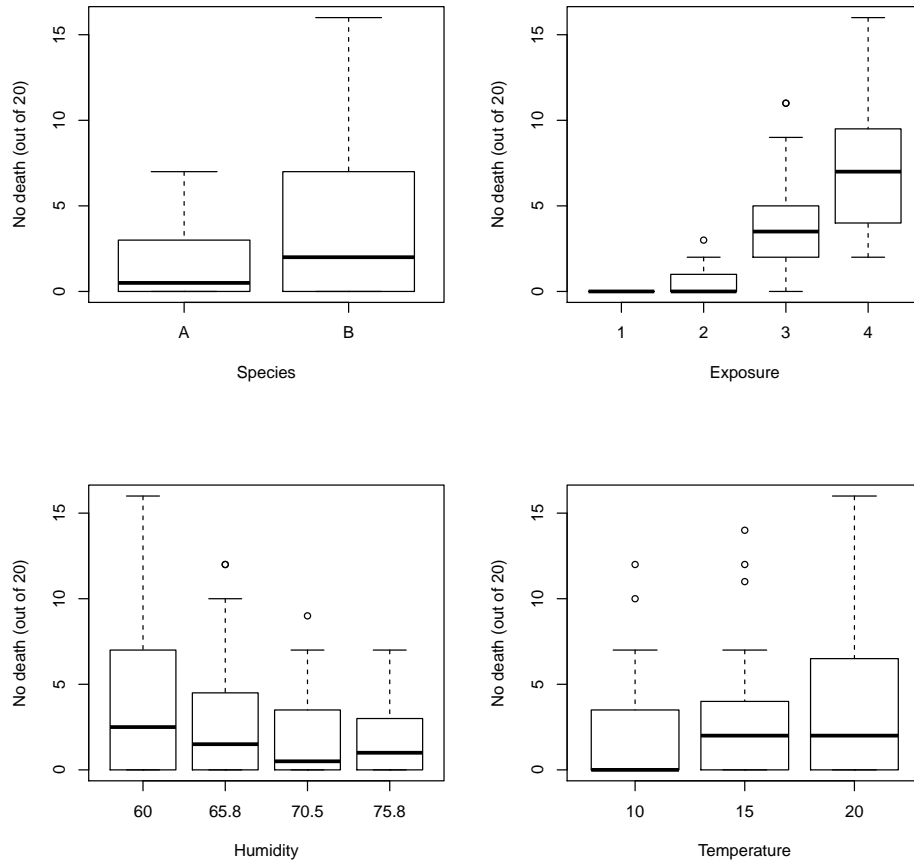
The data set `snails.txt` contains data from an designed experiment studying the survival of snails. Groups of snails of two types, A and B, were held for periods of 1,2,3 or 4 weeks under controlled conditions. Temperature and humidity were kept at predefined levels. The variables are as follows:

<code>species</code>	Snail species, A or B
<code>exposure</code>	Exposure in weeks (1,2,3 or 4)
<code>humidity</code>	Relative humidity (four levels)
<code>temp</code>	Temperature in degrees of Celsius (three levels)
<code>deaths</code>	Number of deaths
<code>n</code>	Number of snails exposed

The main goal is to find whether exposure, humidity, temperature or interactions between these have any effects of the survival probability for snails. You might want start the analysis by investigating whether to include, `exposure`, `humidity` and `temp` as factor variables or numeric variables. When a suitable model has been found perform a residual diagnostic on the model.

Solution

```
> dat<-read.table('snails.txt', head=TRUE)
> par(mfrow=c(2,2))
> boxplot(dat$deaths~dat$species, xlab='Species', ylab='No death (out of 20)')
> boxplot(dat$deaths~dat$exposure, xlab='Exposure', ylab='No death (out of 20)')
> boxplot(dat$deaths~dat$humidity, xlab='Humidity', ylab='No death (out of 20)')
> boxplot(dat$deaths~dat$temp, xlab='Temperature', ylab='No death (out of 20)')
```



```
> dat$resp<-cbind(dat$deaths,dat$n-dat$deaths)
> fit<-glm(formula = resp ~ exposure+I(exposure*exposure)+humidity+temp+species,
+          family = binomial(link = logit),
+          data = dat)
> anova(fit, test='Chisq')
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: resp

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi)
(Intercept)	1	100.00	19	100.00	<2e-16 ***
exposure	3	10.00	16	10.00	0.0001 ***
humidity	3	1.00	13	1.00	0.8000
temp	2	0.50	11	0.50	0.9500
species	1	0.10	10	0.10	0.7500
exposure^2	1	0.01	9	0.01	0.9200

NULL			95	539.72	
exposure	1	332.07	94	207.65	< 2.2e-16 ***
I(exposure * exposure)	1	24.42	93	183.23	7.756e-07 ***
humidity	1	58.17	92	125.06	2.400e-14 ***
temp	1	22.77	91	102.30	1.830e-06 ***
species	1	69.39	90	32.90	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(fit)

Call:

glm(formula = resp ~ exposure + I(exposure * exposure) + humidity +
temp + species, family = binomial(link = logit), data = dat)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7358	-0.4118	-0.1336	0.2021	1.3151

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.73363	1.97628	-3.913	9.11e-05 ***
exposure	5.83709	1.12671	5.181	2.21e-07 ***
I(exposure * exposure)	-0.70173	0.17477	-4.015	5.94e-05 ***
humidity	-0.10549	0.01378	-7.656	1.91e-14 ***
temp	0.09311	0.01916	4.859	1.18e-06 ***
speciesB	1.28802	0.16163	7.969	1.60e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 539.721 on 95 degrees of freedom
Residual deviance: 32.902 on 90 degrees of freedom
AIC: 203.76

Number of Fisher Scoring iterations: 6

> drop1(fit, test='Chisq')

Single term deletions

Model:

```

resp ~ exposure + I(exposure * exposure) + humidity + temp +
      species

```

	Df	Deviance	AIC	LRT	Pr(Chi)	
<none>		32.902	203.76			
exposure	1	79.798	248.66	46.896	7.486e-12	***
I(exposure * exposure)	1	55.070	223.93	22.167	2.499e-06	***
humidity	1	96.484	265.35	63.581	1.539e-15	***
temp	1	57.305	226.17	24.403	7.814e-07	***
species	1	102.296	271.16	69.393	< 2.2e-16	***

```

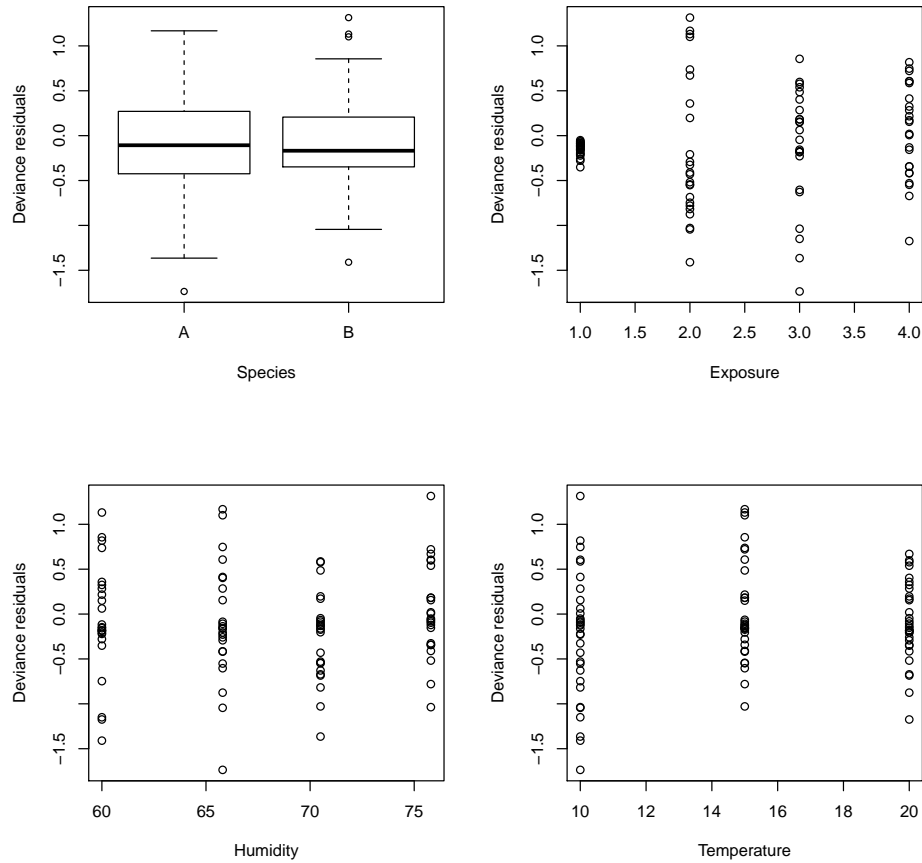
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

> Rd<-residuals(fit,type='deviance')
> par(mfrow=c(2,2))
> plot(dat$species,Rd, xlab='Species', ylab='Deviance residuals')
> plot(dat$exposure,Rd, xlab='Exposure', ylab='Deviance residuals')
> plot(dat$humidity,Rd, xlab='Humidity', ylab='Deviance residuals')
> plot(dat$temp,Rd, xlab='Temperature', ylab='Deviance residuals')

```



```
> Rp<-residuals(fit,type='pearson')
> par(mfrow=c(2,2))
> plot(dat$species,Rp, xlab='Species', ylab='Pearson residuals')
> plot(dat$exposure,Rp, xlab='Exposure', ylab='Pearson residuals')
> plot(dat$humidity,Rp, xlab='Humidity', ylab='Pearson residuals')
> plot(dat$temp,Rp, xlab='Temperature', ylab='Pearson residuals')
```

