

Basic statistical modelling in R GLMs

Clarke van Steenderen
Department of Zoology and Entomology
Rhodes University
South Africa

Clarke.vansteenderen@ru.ac.za



Generalized Linear Models (GLMs)

Linear models assume normality in residuals/error terms → i.e. a constant relationship between mean and variance. But what if this is not met?

We can apply different error structures to a linear model to account for deviations from normality. This means that we can use a GLM, and specify the “family” type depending on the type of data in question

Common families are:

- **Binomial** → two-state data
- **Negative binomial** → count data that is overdispersed
- **Poisson** → count data
- **Gaussian** → normally-distributed; essentially a LM or ANOVA
- **Exponential**

Let's go over an example

Schluter & Smith (1986) had a look at whether different morphological traits of song sparrows had an effect on survival. We'll focus on tarsal length:

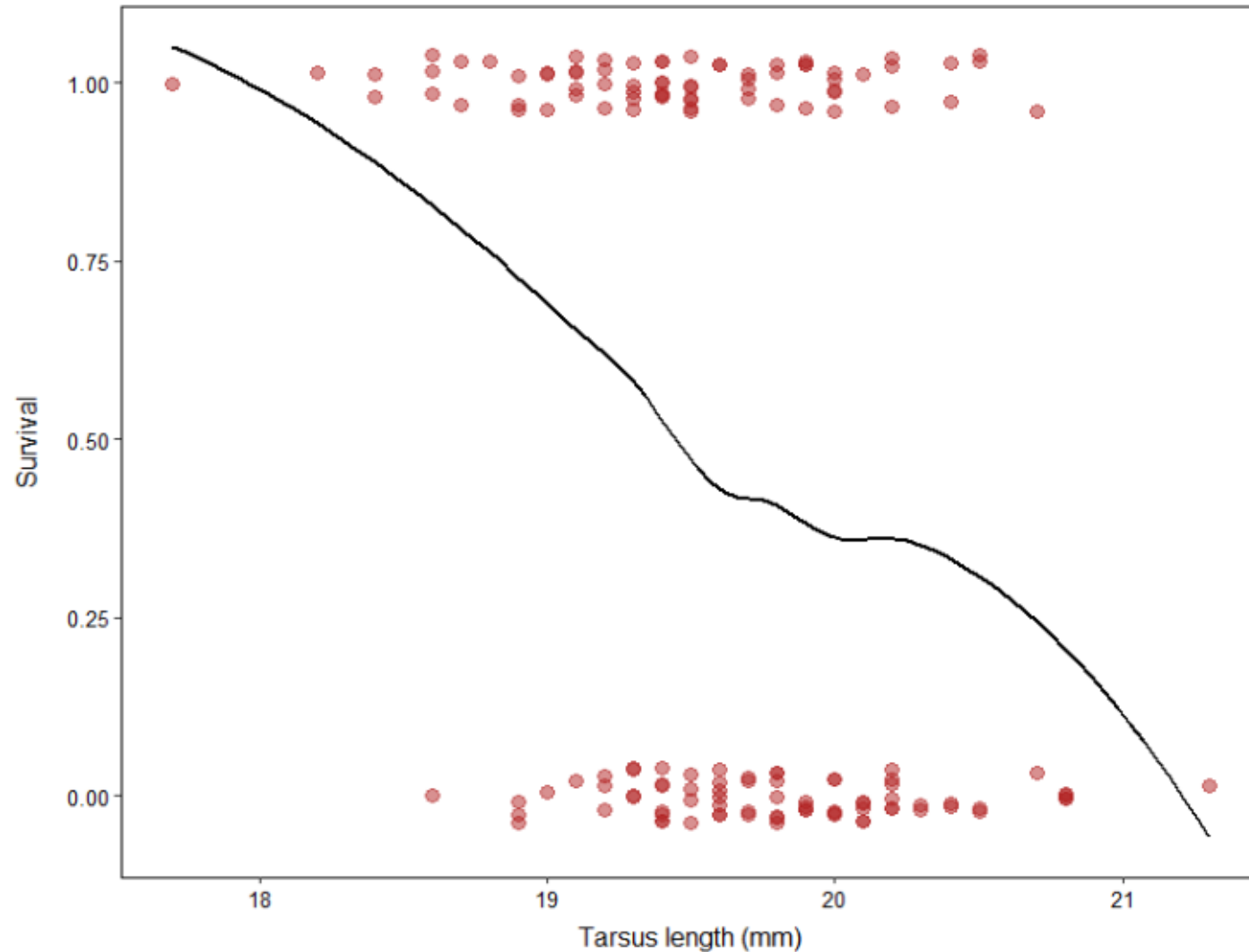
```
> head(sparrow.data)
```

	mass	wing	tarsus	blength	bdepth	bwidth	year	sex	survival	predictions
1	23.7	67.0	17.7	9.1	5.9	6.8	1978	f	1	0.9147634
2	23.1	65.0	19.5	9.5	5.9	7.0	1978	f	0	0.5272762
3	21.8	65.2	19.6	8.7	6.0	6.7	1978	f	0	0.4958586
4	21.7	66.0	18.2	8.4	6.2	6.8	1978	f	1	0.8512375



What relationship do you notice?

Is it good for these sparrows to have small or large feet? Why?



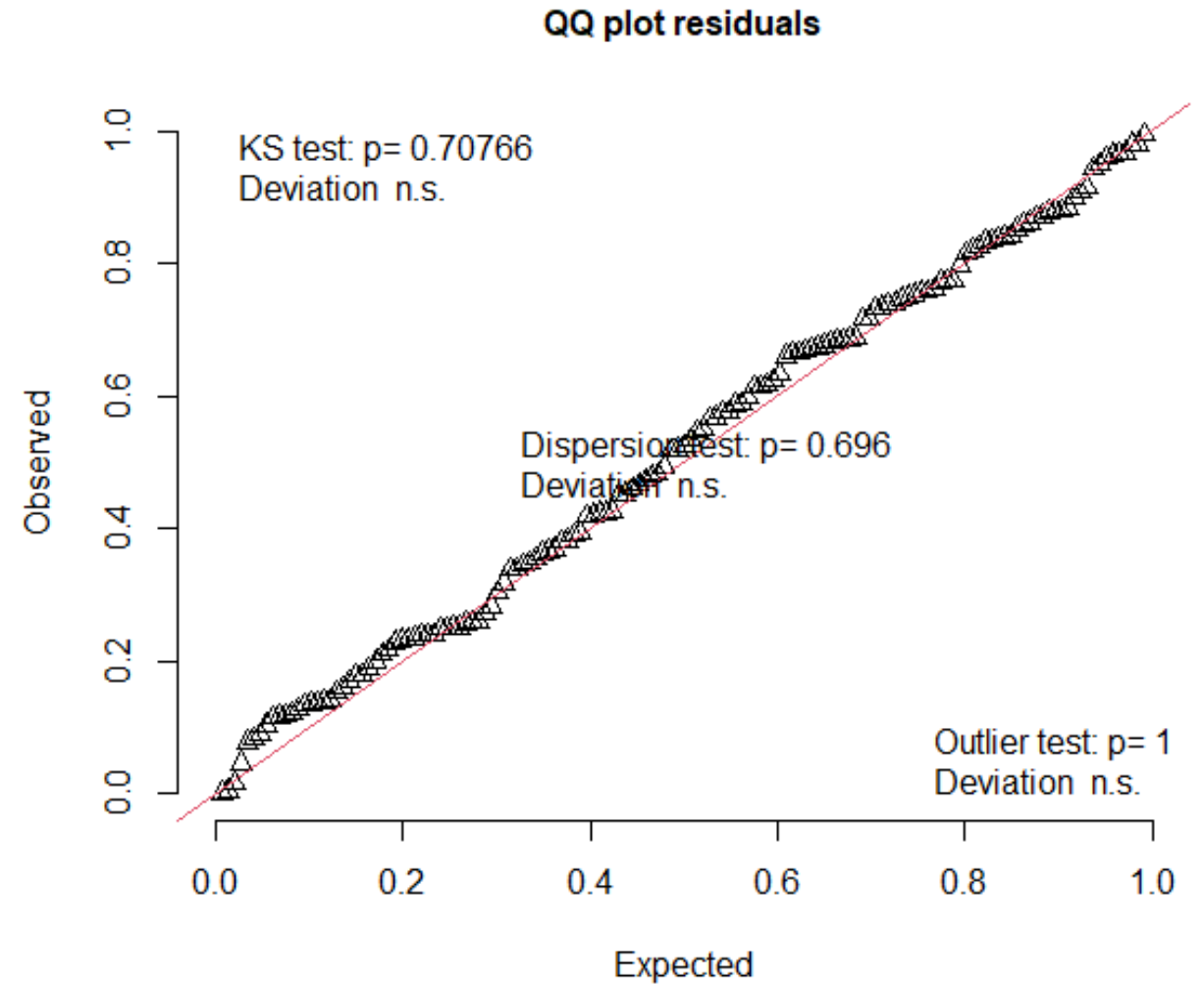
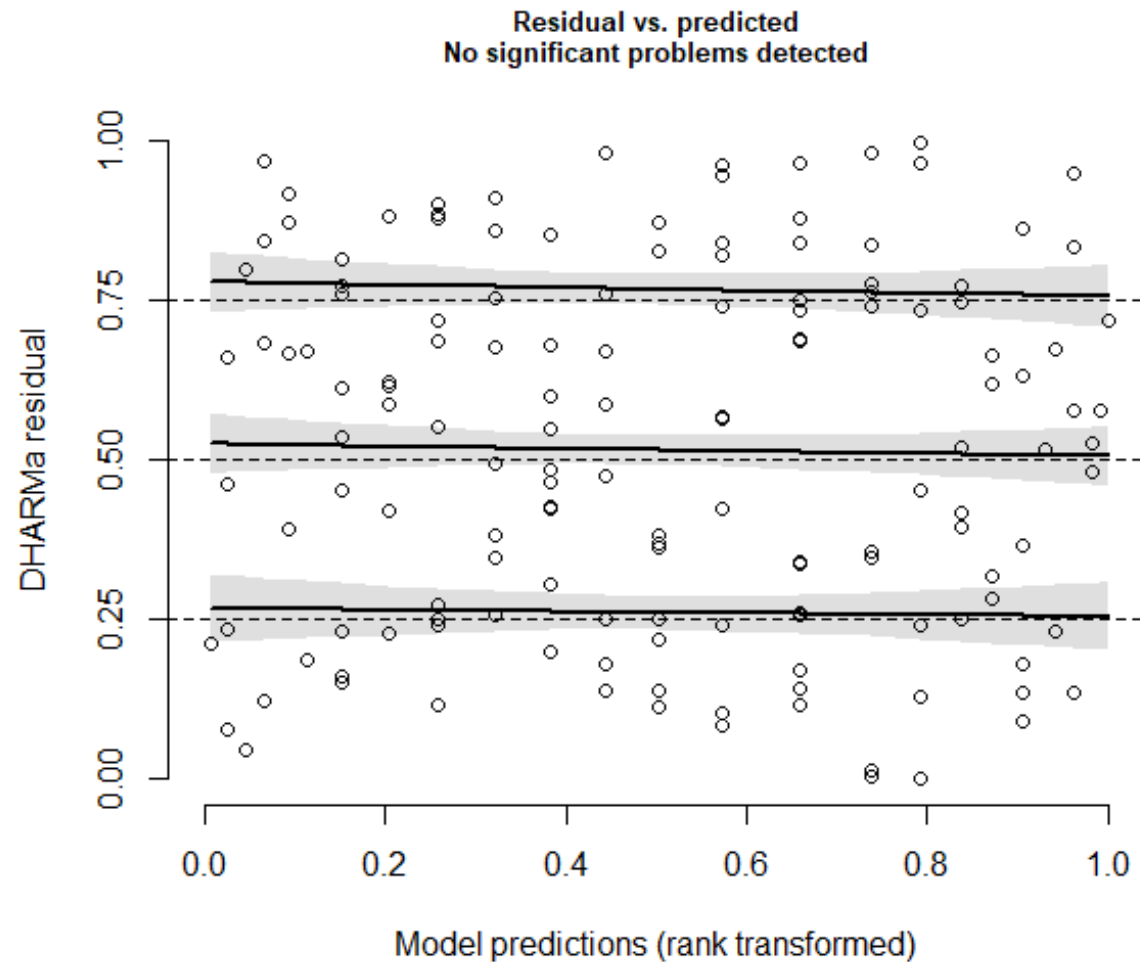
Since we have binomial data (dead or alive), we'll use a binomial GLM using the glmmTMB package

```
sparrow.glm = glmmTMB::glmmTMB(survival ~ tarsus,  
family = binomial(link = "logit"),  
data = sparrow.data)
```

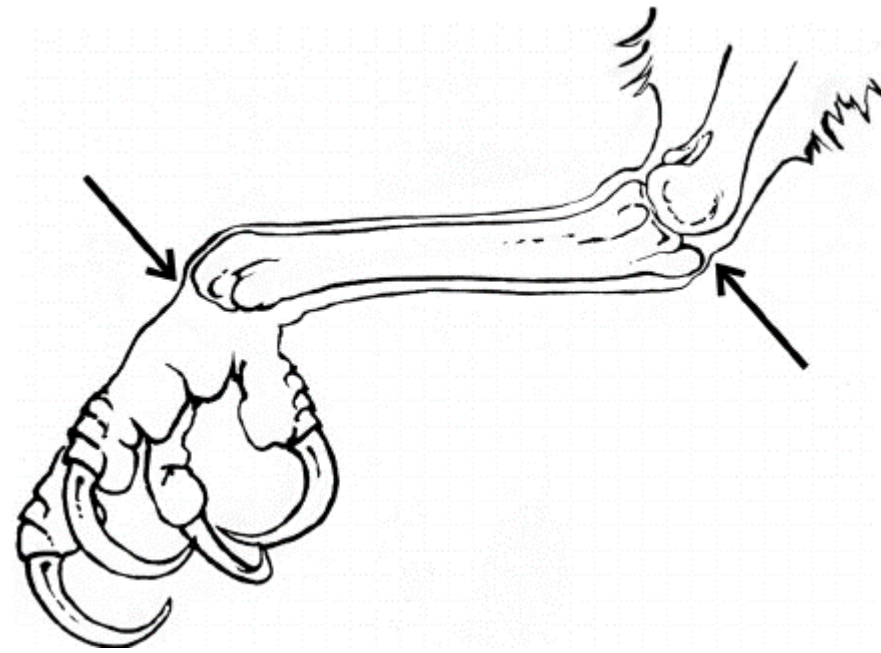
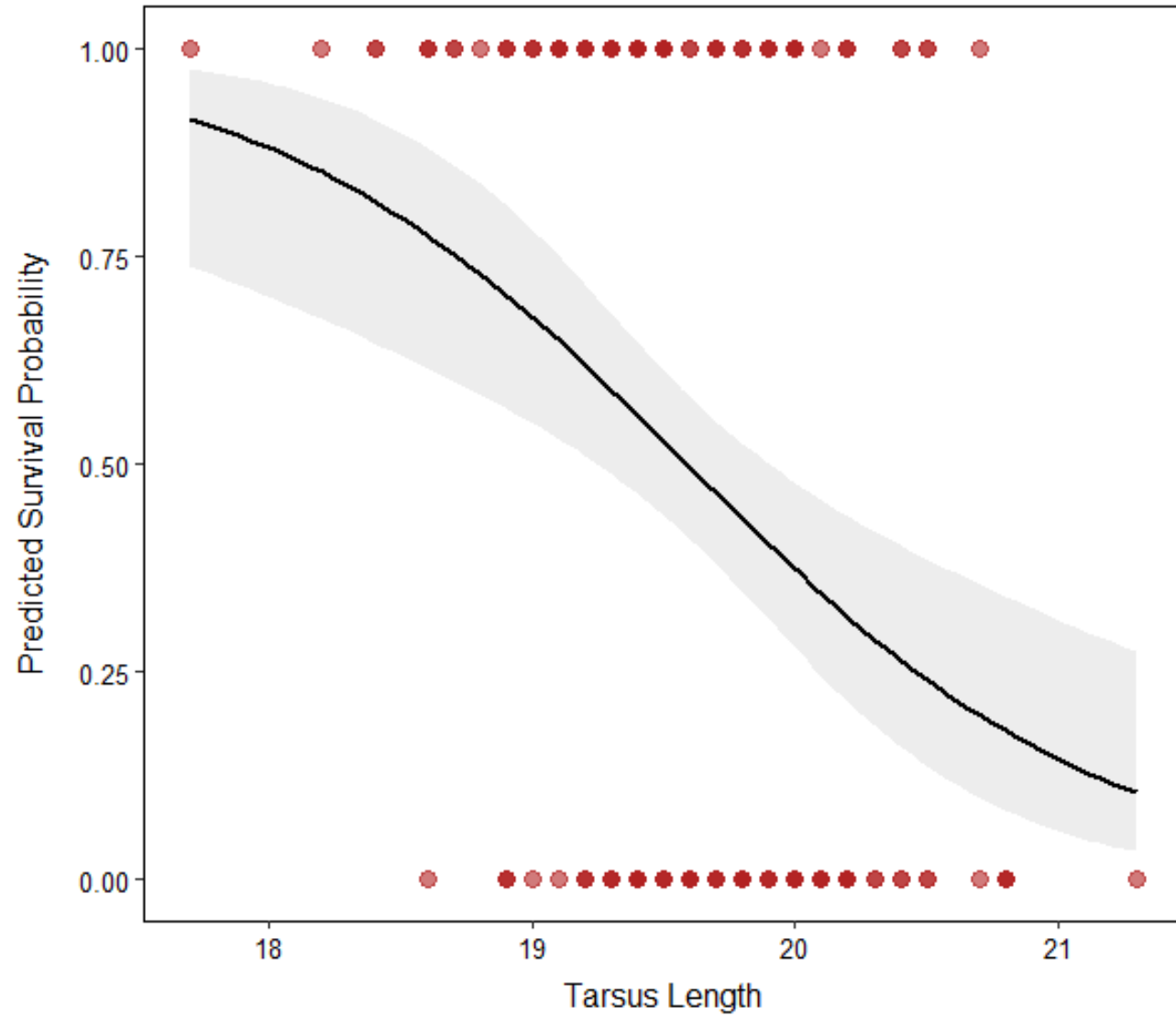
Check the model diagnostics using the DHARMA package

```
DHARMA::plotResiduals(sparrow.glm)  
DHARMA::plotQQunif(sparrow.glm)
```

Looks like a good model 👍



Plot our binomial GLM:



Check for significance using a Wald Chi-square test:

```
> car::Anova(sparrow.glm, test = "Chisq", type = "II")
```

```
Analysis of Deviance Table (Type II Wald chisquare tests)
```

```
Response: survival
```

```
      Chisq Df Pr(>Chisq)  
tarsus 13.391  1  0.0002529 ***
```

```
---
```

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

Tarsus length has a **significant effect** on survival
($\chi^2 = 13.4$, d.f. = 1, $p < 0.001$)

Let's look at the beta coefficients:

```
>summary(sparrow.glm)
```

```
Family: binomial ( logit )  
Formula: survival ~ tarsus  
Data: sparrow.data
```

AIC	BIC	logLik	deviance	df.resid
189.0	195.0	-92.5	185.0	143

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	24.6361 β_0	6.7454	3.652	0.000260 ***
tarsus	-1.2578 β_1	0.3437	-3.659	0.000253 ***

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

Since this was a logistic regression, we need to get the exponentiated value for the beta coefficient (log odds):

$\exp(-1.2578) = 0.28$ (this is now the **odds ratio**. Note the - sign)

For every mm \uparrow in tarsus length, the odds of survival \downarrow by a factor of 0.28

As a percentage: $(0.28 - 1) \times 100 = -72\%$ (i.e. a 72% decrease in the odds of survival)

95% confidence interval:

Lower = $\exp(-1.97) = 0.14$

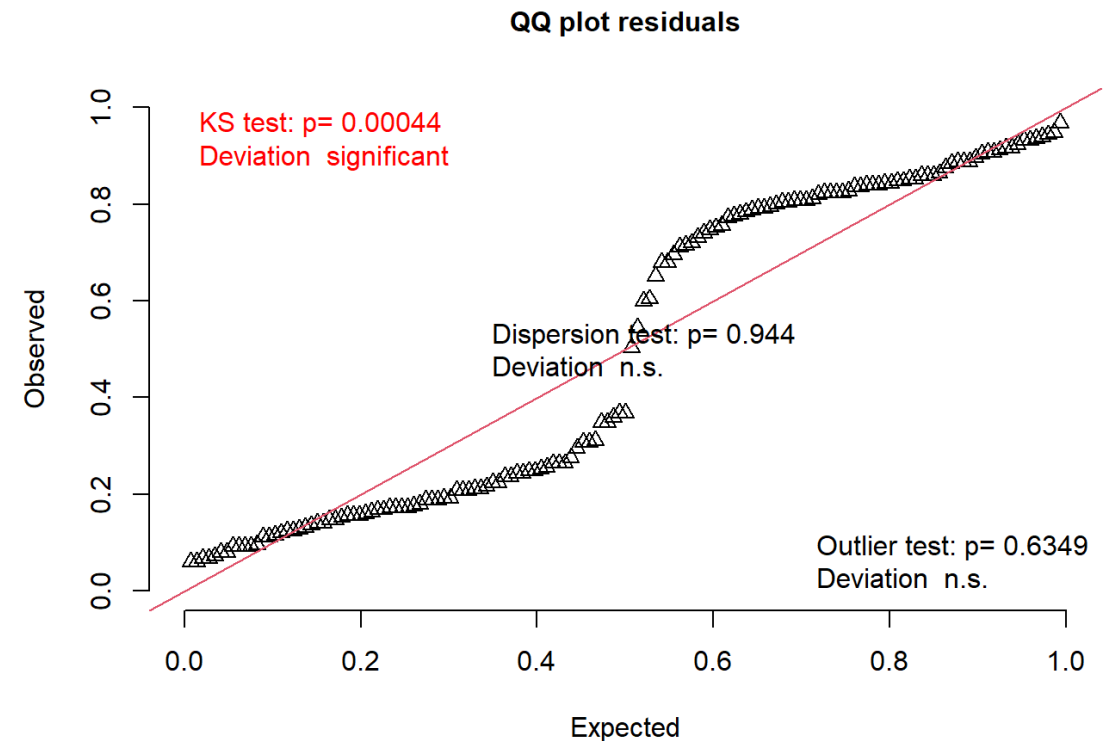
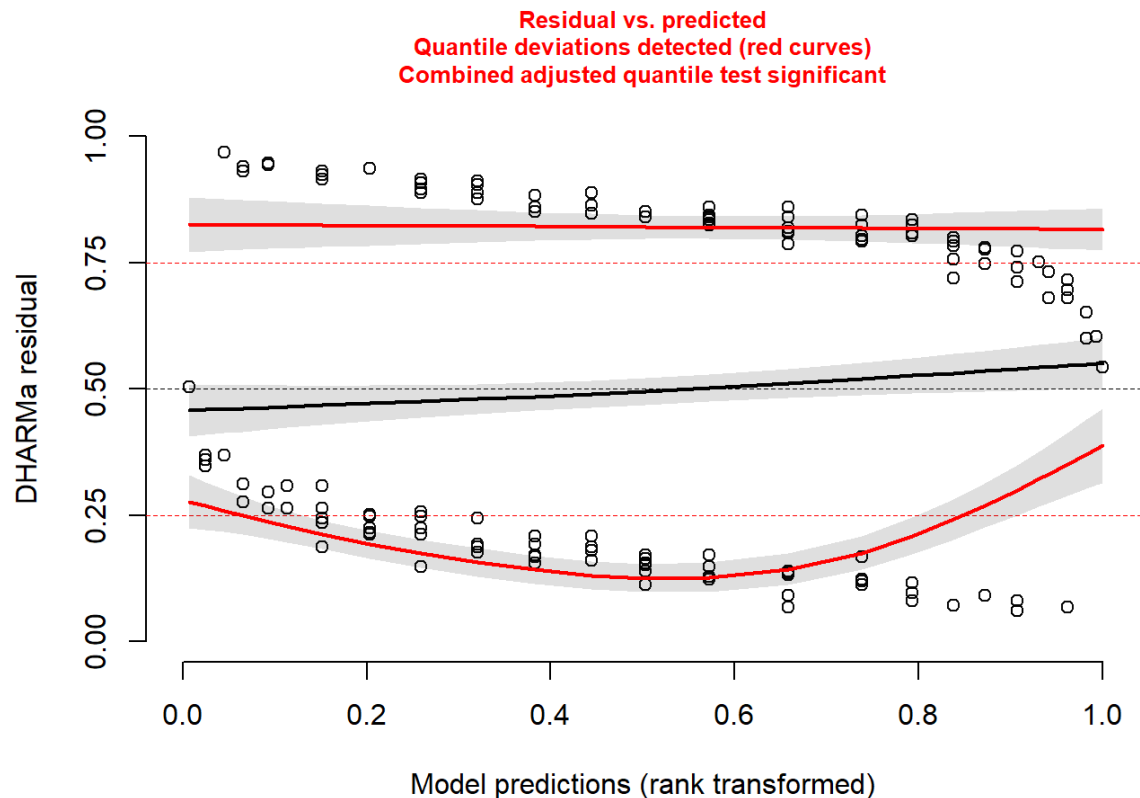
Upper = $\exp(-0.58) = 0.56$

Characteristic	exp(Beta)	95% CI ¹	p-value
tarsus	0.28	0.14, 0.56	<0.001
¹ CI = Confidence Interval			

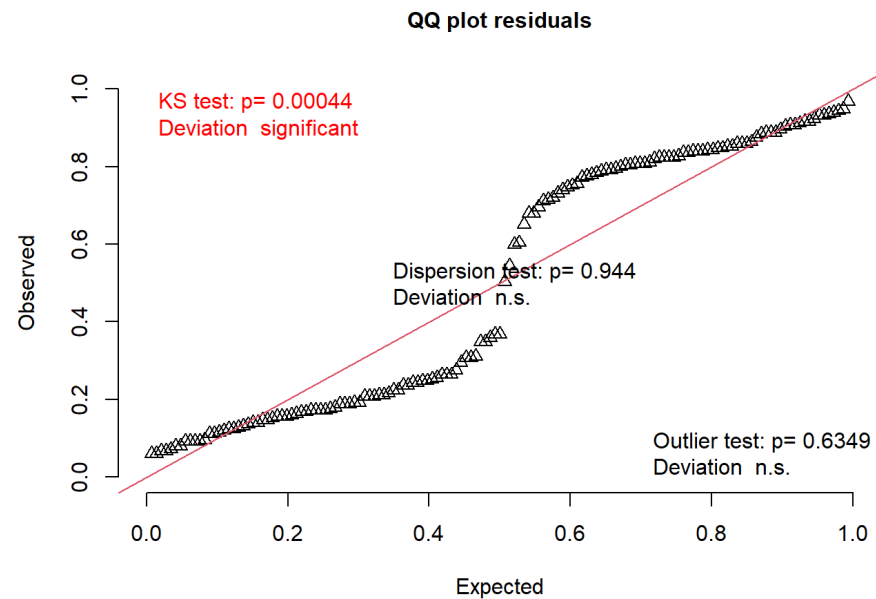
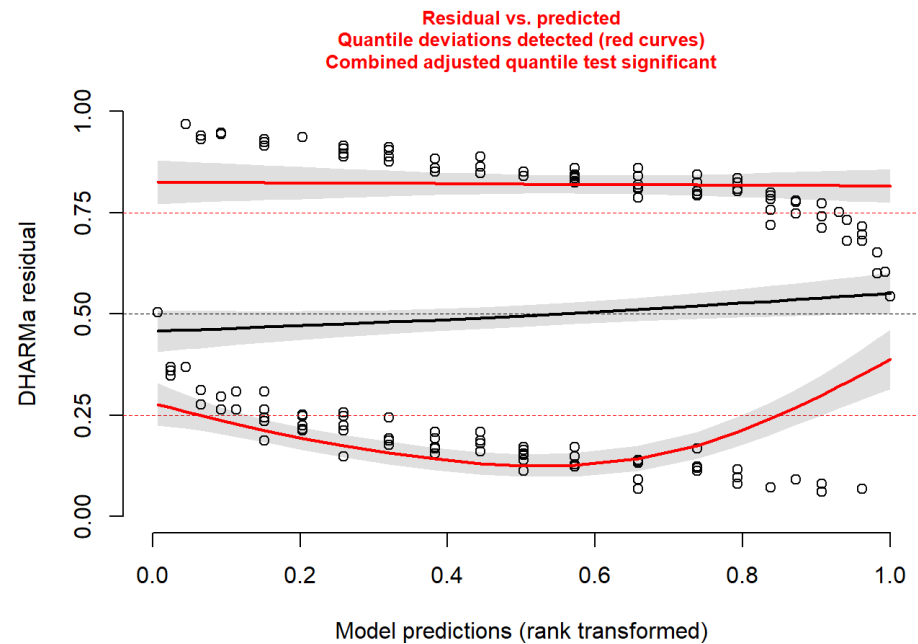
95 out of 100 times, we will get an odds ratio between 0.14 and 0.56

What would happen if we tried running a standard LM, rather than a binomial GLM?

```
sparrow.lm = lm(survival ~ tarsus, data = sparrow.data)  
DHARMA::plotResiduals(sparrow.lm)
```



LINEAR MODEL



BINOMIAL GLM

