

## Problem 6

(a) Download the crabs dataset and show check the first six rows of the dataset.

```
1 # (a) Import package glm2
2 install.packages("glm2")
3 library(glm2)
4
5 # check the first 6 rows of horseshoe crab data
6 head(crabs)
```

(b) Try Identity link Poisson and Identity link Gaussian.

**Identity link Poisson:**

```
1 # (b)
2 # identity link poisson
3 fit1 <- glm2(Satellites ~ Width + factor(Dark) + factor(GoodSpine) + Rep1 +
4   Rep2, data = crabs, family = poisson(link = "identity"))
5 summary(fit1)
```

The result is presented by `summary(fit1)`:

```
> summary(fit1)

Call:
glm2(formula = Satellites ~ Width + factor(Dark) + factor(GoodSpine) +
      Rep1 + Rep2, family = poisson(link = "identity"), data = crabs,
      start = rep(1, 6))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0617  -1.9284  -0.4670   0.9494   4.3603

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -9.30242    1.49377  -6.227 4.74e-10 ***
width           0.49185    0.05590   8.799 < 2e-16 ***
factor(Dark)yes -0.61294    0.26242  -2.336  0.01951 *
factor(GoodSpine)yes -0.10751    0.26855  -0.400  0.68890
Rep1            0.04800    0.01660   2.892  0.00383 **
Rep2           -0.04939    0.01690  -2.922  0.00347 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 545.09  on 167  degrees of freedom
AIC: 912.39

Number of Fisher Scoring iterations: 22
```

**Identity link Gaussian:**

```

1 # identity link gaussian
2 fit2 <- glm2(Satellites ~ Width + factor(Dark) + factor(GoodSpine) + Rep1 +
  Rep2, data = crabs, family = gaussian(link = "identity"))
3 summary(fit2)

```

The result is presented by **summary(fit2)**:

```

> summary(fit2)

Call:
glm2(formula = Satellites ~ Width + factor(Dark) + factor(GoodSpine) +
  Rep1 + Rep2, family = gaussian(link = "identity"), data = crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.3907  -2.2628  -0.5002   1.7883  10.9386

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -8.62002    3.00977  -2.864  0.00472 **
width           0.47947    0.11153   4.299 2.91e-05 ***
factor(Dark)yes -0.59235    0.50258  -1.179  0.24023
factor(GoodSpine)yes 0.04635    0.51554   0.090  0.92847
Rep1            0.05258    0.03536   1.487  0.13896
Rep2           -0.05814    0.03657  -1.590  0.11376
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 8.761317)

    Null deviance: 1704.9  on 172  degrees of freedom
Residual deviance: 1463.1  on 167  degrees of freedom
AIC: 874.32

Number of Fisher Scoring iterations: 2

```

(c) Compare the deviances and convergence of two models:

```

1 # (c) Compare two models
2 noquote(c("deviances:", fit1$dev, fit2$dev))
3 noquote(c("converged:", fit1$conv, fit2$conv))

```

The result is as follows:

```

> # (c) Compare two models
> noquote(c("deviances:", fit1$dev, fit2$dev))
[1] deviances:      545.093892666259 1463.13991504975
> noquote(c("converged:", fit1$conv, fit2$conv))
[1] converged:  TRUE      TRUE

```

**Interpretation:** Both models are convergent and the deviance of identity link to poisson model is far less than identity link to gaussian model.

(d) Conduct the likelihood ratio test regarding null hypothesis,  $H_0$ : No colour effect. Compute the model using identity link to poisson without 'Dark', which represents color term. And use function **lrtest** to conduct likelihood ratio test. The code and the test result are as follows:

```

1 # (d) likelihood test
2 library(lmtest)
3 fit1_null <- glm2(Satellites ~ Width + factor(GoodSpine) + Rep1 + Rep2,
4                   data = crabs, family = poisson(link = "identity"), start =
5                   rep(1,5))
6 lrtest(fit1_null, fit1)

```

Results:

```

> lrtest(fit1_null, fit1)
Likelihood ratio test

Model 1: Satellites ~ Width + factor(GoodSpine) + Rep1 + Rep2
Model 2: Satellites ~ Width + factor(Dark) + factor(GoodSpine) + Rep1 +
Rep2
#Df  LogLik Df  Chisq Pr(>Chisq)
1    5 -452.74
2    6 -450.20  1 5.0794    0.02421 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Hence, we can reject the null hypothesis at 95% confidence.