Error terms in generalized linear models

Xulong Wang

November 5, 2015

Model structure: linear model

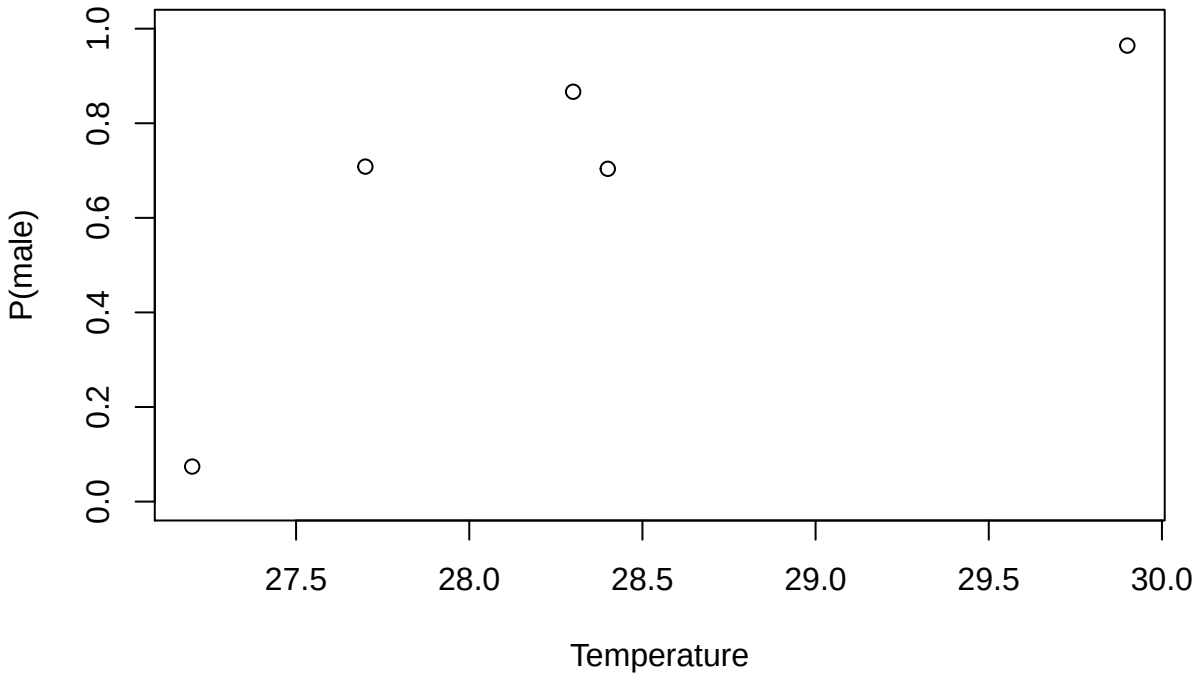
$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

Model structure: generalized linear model (logistic)

$$Prob(Y_i = 1) = \pi_i$$
$$E(Y_i|X_i) = \pi_i = \beta_0 + \beta_1 X_i$$
$$Var(Y_i) = \pi_i(1 - \pi_i)$$

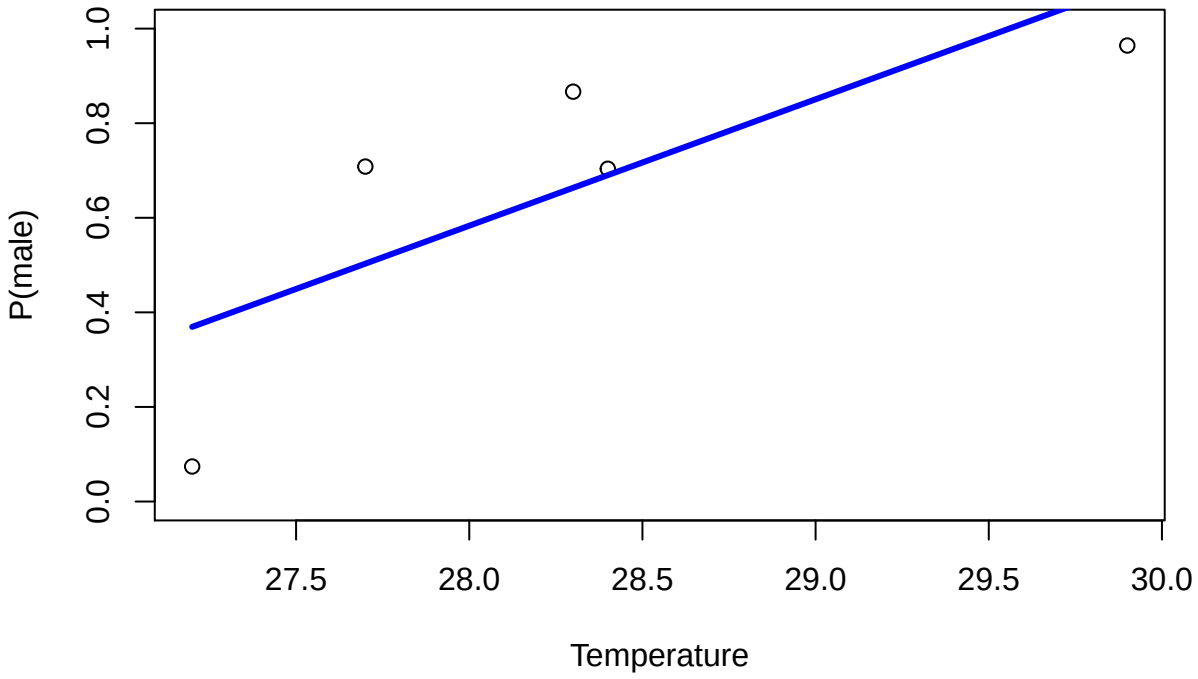
Example: water temperature can predict male turtles proportion

```
library(gtools)

rm(list = ls())

temp = c(27.2, 27.7, 28.3, 28.4, 29.9)
nmale = c(2, 17, 26, 19, 27)
nfemale = c(25, 7, 4, 8, 1)
ntur = nmale + nfemale
pmale = nmale / ntur

plot(temp, pmale, ylim = c(0, 1), xlab = "Temperature", ylab = "P(male)")
```



OLS directly on probability

```
lm0 = lm(pmale ~ temp)
res = residuals(lm0)

plot(temp, pmale, ylim = c(0, 1), xlab = "Temperature", ylab = "P(male)")
lines(temp, lm0$fitted, col = "blue", lwd = 3)
```



OLS on logit-transformed probability

```
log.pmale = logit(pmale)
lm1 = lm(log.pmale ~ temp)
summary(lm1)
```

```
##
## Call:
## lm(formula = log.pmale ~ temp)
##
## Residuals:
##       1       2       3       4       5
## -1.3837  1.1107  0.9930 -0.1976 -0.5224
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -51.1122    16.9415  -3.017   0.0569 .
## temp          1.8371     0.5983   3.070   0.0545 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.217 on 3 degrees of freedom
## Multiple R-squared:  0.7586, Adjusted R-squared:  0.6781
## F-statistic: 9.428 on 1 and 3 DF,  p-value: 0.05454
```

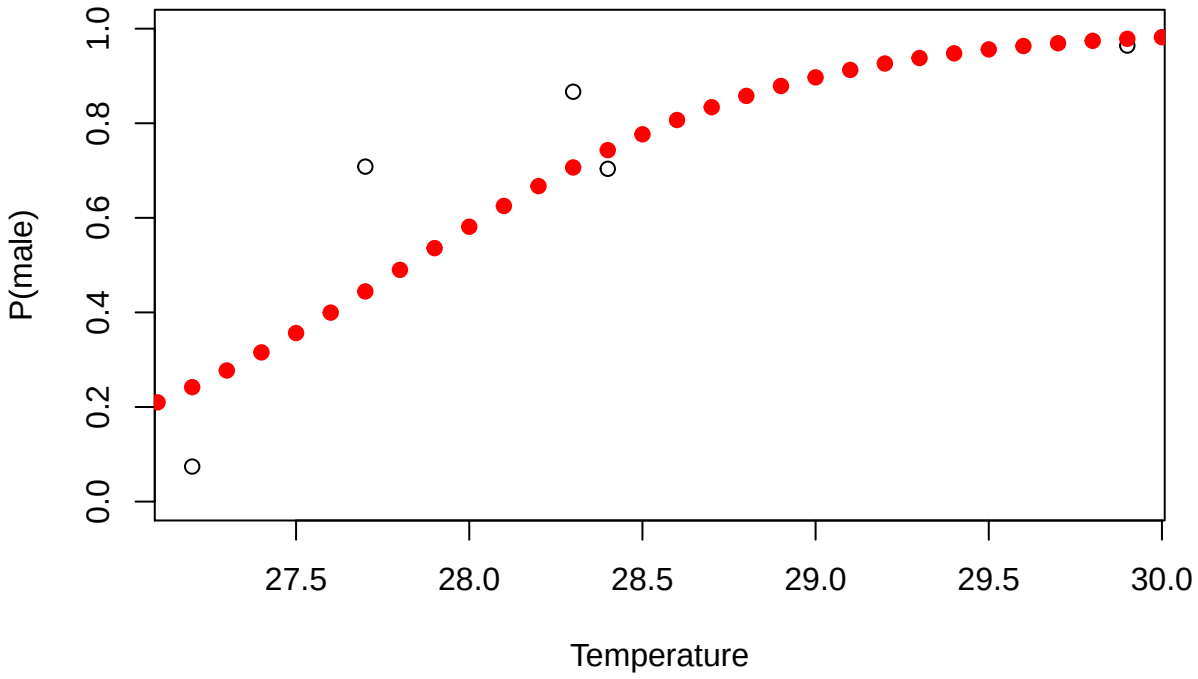
```
fitted(lm1)
```

```
##          1          2          3          4          5
## -1.1420115 -0.2234416  0.8788422  1.0625562  3.8182657
```

```
inv.logit(fitted(lm1))
```

```
##         1         2         3         4         5
## 0.2419512 0.4443709 0.7065822 0.7431787 0.9785063
```

```
x = seq(27, 30, 0.1)
y = inv.logit(summary(lm1)$coefficients[1] + x * summary(lm1)$coefficients[2])

plot(temp, pmale, ylim = c(0, 1), xlab = "Temperature", ylab = "P(male)")
points(x, y, col = "red", pch = 19)
```

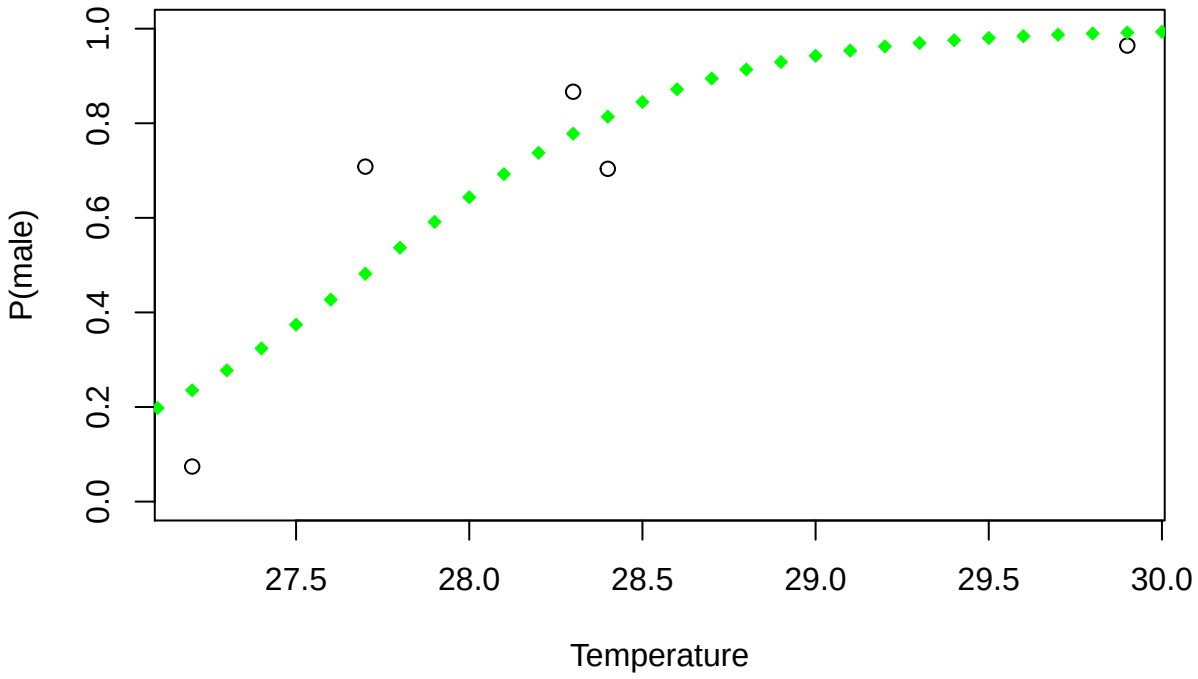


MLE

```
glmf = glm(pmale ~ temp, family = binomial, weight = ntur)
fitted(glmf)
```

```
##         1         2         3         4         5
## 0.2353639 0.4818202 0.7779686 0.8138117 0.9917685
```

```
y = inv.logit(summary(glmf)$coefficients[1] + x * summary(glmf)$coefficients[2])

plot(temp, pmale, ylim = c(0, 1), xlab = "Temperature", ylab = "P(male)")
points(x, y, col = "green", pch = 18)
```



MLE and OLS comparison

```
sum(dbinom(nmale, ntur, fitted(glmf), log = T))
```

```
## [1] -14.77124
```

```
sum(dbinom(nmale, ntur, inv.logit(fitted(lm1)), log = T))
```

```
## [1] -15.76246
```

```
x = 1:10
y = 2:11 + rnorm(10, 0, 1)

lm(y ~ x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##      1.7239       0.8834
```

```
glm(y ~ x)
```

```
##
## Call:  glm(formula = y ~ x)
```

```
##
## Coefficients:
## (Intercept)            x
##      1.7239       0.8834
##
## Degrees of Freedom: 9 Total (i.e. Null);  8 Residual
## Null Deviance:       68.74
## Residual Deviance: 4.351     AIC: 26.06
```