ES/STT 7140: Homework 6

Brandon M. Greenwell

Wednesday, April 4, 2018 (due Wednesday, April 11, 2018)

Question 1 TBD.

a)

```
# Load the data
url <- paste0("http://users.stat.ufl.edu/~presnell/Courses/",
              "sta4504-2000sp/R/Data/crabs.dat")
crabs <- read.table(url, header = TRUE)
head(crabs)  # print first few observations

# Scatterplot of raw data
plot(
  formula = satell ~ weight,
  data = crabs,
  xlab = "Width (cm)",
  ylab = "Number of satellites"
)
```

b) For a female crab, let μ be the expected number of satellites and $x = width$. Using the R code below, fit a Poisson loglinear model of the form

$$\log(\mu) = \beta_0 + \beta_1 x$$

. For a one-unit increase in width (e.g., one centimeter), what is the multiplicative effect on $\widehat{\mu}$ for a 1-cm increase in width?

```
# Fit a Poisson loglinear model
fit <- glm(satell ~ width, data = crabs, family = poisson(link = "log"))
summary(fit)  # print model summary
```

Question 2 The R data frame `ldeaths` contains monthly death rates from three lung diseases in the UK over a period of several years (see `?ldeaths` for details and a reference). One possible model for the data is that they can be treated as Poisson random variables with a seasonal component and a long term trend component, as follows:

$$E(deaths) = \beta_0 + \beta_1 t + \alpha \sin(2\pi \times toy/12 + \phi),$$

where β_0, β_1, α, and ϕ are parameters, t is time since the start of the data, and toy is time of year, in months (January being month 1). The data can be loaded into R using the following snippet of code:

```
death <- as.numeric(ldeaths)
month <- rep(1:12, 6)
time <- 1:72
```

By making use of properties of sines and cosines (i.e., basic trigonometry), we can get this model into a form suitable for fitting using `glm()`, and fit it using the following code:

```
fit <- glm(death ~ sin(month/12*2*pi) + cos(month/12*2*pi),
           family = poisson(link = "identity"))
summary(fit)
```

Use the following code to plot the raw data time series on a plot, with the predicted time series overlaid. Additionally, the code also constructs a scatterplot of the (deviance) residuals versus the fitted values and a

Q-Q plot of the residuals. Based on these plots alone, does the model appear to provide an adequate fit to the data? Write your answer using a paragraph (i.e., 5 – 7 complete sentences).

```
# Plot fitted mean response
plot(time, death, type = "l")
lines(time, fitted(fit), col = 2)

# Residual plots
par(mfrow = c(1, 2))
plot(fit, which = 1:2)
```