Biostats VHM 801/802/831 Courses
Summer 2006, Atlantic Veterinary College, PEI
Henrik Stryhn

Notes on Regression Modelling of Periodic (Seasonal) Effects

The purpose of these notes is to outline how to set up a model with periodic (or seasonal, e.g. yearly recurring) patterns of sine/cosine wave shape. This can be done within the framework of a linear model (or any of its extensions, such as modelling on linear predictor scale in a generalized linear model), provided that the period of the sine wave is considered as fixed and not to be estimated from the data. For example, when focusing on yearly recurring patterns, the frequency of the sine wave has been fixed to give a period of one year (see below). However, some formulae are necessary to translate the coefficients from the linear equation to the primary parameters of the sine wave: the amplitude and phase shift.

1 The basic sine wave model

The general equation of a sine wave as a function of time t is the following:

$$f(t) = \mu + \alpha \sin(\omega t + \phi) = \mu + \alpha \sin(\omega(t + \phi_0)). \tag{1}$$

The parameters have the interpretations:

- μ is the average level across one period of the sine wave,
- α is the amplitude, or half the largest difference between values taken by the sine wave ("maximum – minimum"),
- ω is the frequency, or 2π divided by the period of the sine wave; if t is measured in days, one should take $\omega = 2\pi/365$ to obtain a yearly period of the sine wave (where π is the mathematical constant, $\pi = 3.14159265\ldots$),
- ϕ is the phase shift controlling where the sine wave peaks:
  - $f(t) = 0$ when $\omega t + \phi = k\pi$ ($k = \ldots, -2, -1, 0, 1, 2, \ldots$), that is, $t = (k\pi - \phi)/\omega$,
  - $f(t)$ = maximal when $\omega t + \phi = \pi/2 + 2k\pi$, that is, $t = (\pi/2 + 2k\pi - \phi)/\omega$, e.g., $t = (\pi/2 - \phi)/\omega$ and $t = (5\pi/2 - \phi)/\omega$,
  - $f(t)$ = minimal when $\omega t + \phi = -\pi/2 + 2k\pi$, that is, $t = (-\pi/2 + 2k\pi - \phi)/\omega$, e.g., $t = (-\pi/2 - \phi)/\omega$ and $t = (3\pi/2 - \phi)/\omega$,
- $\phi_0 = \phi/\omega$ is the phase shift measured on the same scale as t.

1.1 Example: Bulk milk somatic cell counts (SCCs) during one year in a herd

We consider a small part of a dataset collected from 300 Dutch herds in the years 1992–1995 ([1]). The outcome of interest is the (natural) logarithmic somatic cell count in the milk samples; Figure 1 (next page) shows the observed values as well as the fitted sine wave curve, with its parameters. The estimates are:

$$\begin{aligned} \hat{\mu} &= 4.3655, & \mathrm{SE}(\hat{\mu}) &= 0.0451, \\ \hat{\alpha} &= -0.5847, & \mathrm{SE}(\hat{\alpha}) &= 0.0651, \\ \hat{\phi} &= 0.3180, & \mathrm{SE}(\hat{\phi}) &= 0.1063. \end{aligned}$$

Converted to days, the phase shift is $\hat{\phi}_0 = 365 \cdot \hat{\phi}/2\pi = 18.5$ days, corresponding to a peak at day $0.75 \cdot 365 - 18.5 = 255$.

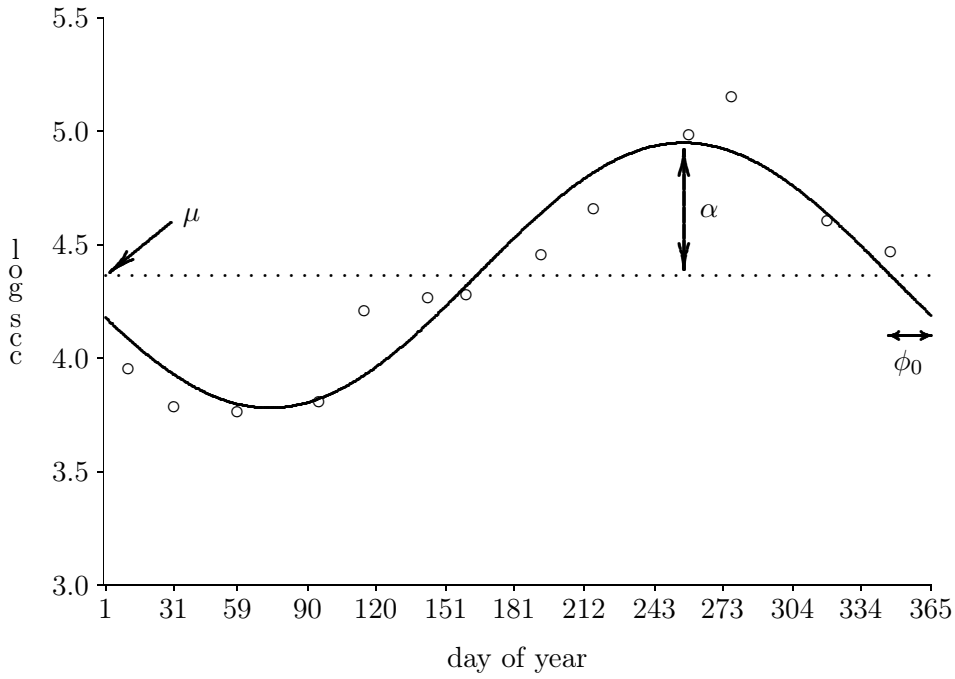

Figure 1: Bulk milk log SCC in a single herd during one year (days 1–365), with overlaid fitted sine wave curve.

2 The linear sine wave model

Equation (1) is linear in the parameters μ and α but non-linear in the parameter ϕ. However, by utilizing standard trigonometric formulae the *same* model can be parametrized linearly in all its parameters:

$$f(t) = \mu + \beta_1 \sin(\omega t) + \beta_2 \cos(\omega t). \tag{2}$$

Therefore, Equation (2) leads to a linear (possibly extended or generalized) statistical model which considerably facilitates the statistical analysis. In this model, μ is unchanged from Equation (1); however, the parameters β_1 and β_2 have no intuitive interpretation. It is therefore often natural to convert them to the parameters α and ϕ by the formulae:[1]

$$\phi = \begin{cases} \arctan(\beta_2/\beta_1) & \text{for } \beta_1 \neq 0, \\ \pi/2 & \text{for } \beta_1 = 0, \end{cases} \tag{3}$$

$$\alpha = \beta_1/\cos(\phi) = \text{sign}(\beta_1)\sqrt{\beta_1^2 + \beta_2^2}, \tag{4}$$

where arctan denotes the inverse tangent function, and $\text{sign}(x) = 1$ for $x \geq 0$ and $= -1$ for $x < 0$. These calculations are illustrated by the somatic cell count data below.

One problem arising from the indirect estimation of α and ϕ through the linear equation (2) is the lack of standard errors for $\hat{\alpha}$ and $\hat{\phi}$. The traditional solution to use the "delta method" (e.g., Section 6.1.2 of [2]) — an approximation formula based on the estimated variances s_1^2 and s_2^2 for $\hat{\beta}_1$ and $\hat{\beta}_2$,

[1] Applying the "addition formula": $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$, to the right hand side of (1) and equating it to the right hand side of (2) yields the relations: $\beta_1 = \alpha\cos(\phi)$ and $\beta_2 = \alpha\sin(\phi)$, from which the formulae (3) and (4) follow immediately. The second form of (4) follows after some rewriting from the trigonometric relation: $\cos(x) = 1/\sqrt{1 + \tan^2(x)}$.

respectively, as well as the estimated covariance s_{12} between these estimates. The resulting formulae are:[2]

$$\mathrm{SE}(\hat{\phi})^2 \approx (s_1^2\hat{\beta}_2^2 + s_2^2\hat{\beta}_1^2 - 2s_{12}\hat{\beta}_1\hat{\beta}_2)/(\hat{\beta}_1^2 + \hat{\beta}_2^2)^2, \tag{5}$$

and

$$\mathrm{SE}(\hat{\alpha})^2 \approx (s_1^2\hat{\beta}_1^2 + s_2^2\hat{\beta}_2^2 + 2s_{12}\hat{\beta}_1\hat{\beta}_2)/(\hat{\beta}_1^2 + \hat{\beta}_2^2). \tag{6}$$

2.1 Example: Bulk milk somatic cell counts (cont)

A linear model based on Equation (2) gave the estimates:

$$\begin{aligned} \hat{\beta}_1 &= -0.55537, \quad \mathrm{SE}(\hat{\beta}_1) = 0.0650 \quad (\text{or } s_1^2 = 0.0042256), \\ \hat{\beta}_2 &= -0.18283, \quad \mathrm{SE}(\hat{\beta}_2) = 0.0623 \quad (\text{or } s_2^2 = 0.0038813), \end{aligned}$$

as well as the estimated covariance $s_{12} = 0.0000883$ (we note in passing that since this value is much smaller than the two variances, the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are almost uncorrelated). We can now apply the formulae (3) and (4) to recompute the previously given values for $\hat{\phi}$ and $\hat{\alpha}$,

$$\begin{aligned} \hat{\phi} &= \arctan((-0.18283)/(-0.55537)) = \arctan(0.329204) = 0.3180, \\ \hat{\alpha} &= (-0.55537)/\cos(0.3180) = -0.5847 = -\sqrt{(-0.55537)^2 + (-0.18283)^2}. \end{aligned}$$

Similar calculations by the approximation formulae (5) and (6) lead to the same values for the respective standard errors as previously given, without any noticeable approximation error.

References

[1] Olde Riekerink, R., Stryhn, H. & Barkema, H. W. (2006), The effect of season on somatic cell count and the incidence of clinical mastitis. Submitted manuscript.

[2] Weisberg, S. (2005). *Applied Linear Regression*, 3rd ed. Wiley.

[2] The delta formula for a general function $g(x, y)$ is: $\mathrm{Var}(g(x,y)) \approx \mathrm{Var}(x)(\partial g/\partial x)^2 + \mathrm{Var}(y)(\partial g/\partial y)^2 + 2\,\mathrm{Cov}(x,y)(\partial g/\partial x)(\partial g/\partial y)$. We use this formula twice, for $g(x,y) = \arctan(y/x)$ and for $g(x,y) = \sqrt{x^2+y^2}$.