Sinusoidal Regression

Owen Ward

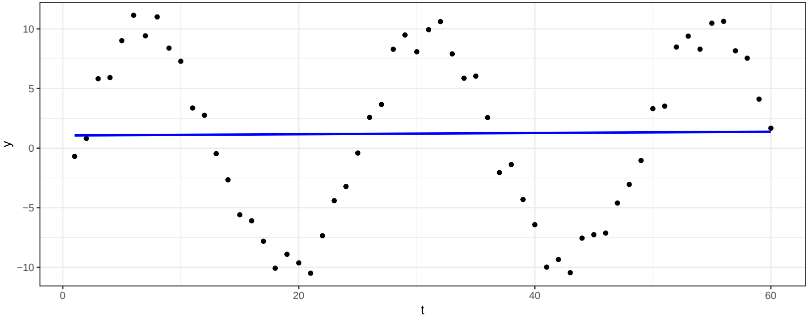
11/7/2018

Scientists often need to adapt regression to allow for cyclical or periodic relationships

- Linear regression is a simple (but powerful) tool for quantifying the relationship between two variables.
- However, it only measures linear association—one of many possible relationships.
- Time series are generally periodic; the same pattern repeats at certain time points.
  - e.g. traffic is likely to increase during rush hour and decrease at night
  - e.g. sales of air conditioners are likely to increase every summer, then decrease in the winter
- A cursory linear regression will not capture these relationships.

Linear regression doesn't automatically work.

```
library("tidyverse"); library("knitr")
theme_set(theme_bw())
sim <- tibble(t = 1:60,
   y = 10 * sin(2*pi*t/24+6) + rnorm(60))
ggplot(sim, aes(t,y)) +
  geom_point() +
  geom_smooth(method = "lm",color = "blue",se = FALSE)
```



Linear regression can be augmented to account for periodicity

Given some knowledge of trigonometry, the previous curve looks something like a sinusoid (Sine of Cosine function). One way to capture this relationship is to fit the data with the Sine curve

$$Y_t = A \sin(2\pi\omega t + \phi) + B.$$

A, ω, ϕ and B all have intuitive interpretations, but this formula—as it is currently written—is still not in the form of linear regression.

To do that, use the trigonometric identity:

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

Who knew trigonometric identities were useful?

Combining the two formulas, we get

$$Y_t = A\cos\phi\sin(2\pi\omega t) + A\sin\phi\cos(2\pi\omega t) + B.$$

Letting

$$X_1 = \sin(2\pi\omega t), \quad X_2 = \cos(2\pi\omega t),$$
$$\alpha_1 = A\cos\phi, \quad \alpha_2 = A\sin\phi,$$

then we have

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + B.$$

This is a linear regression model in the new variables X_1, X_2.

New Linear Regression

- Running a linear regression on these new variables will give us estimates $\hat{\alpha}_1, \hat{\alpha}_2, \hat{B}$.
- We want to use these to get estimates of A, ϕ and ω (B is the same in both parameterizations).
- We can again use trigonometry to get back to the original problem.

Going back to Sinusoid

- We have

$$\alpha_1^2 + \alpha_2^2 = A^2 \left(\cos^2\phi + \sin^2\phi\right) = A^2,$$

  so $A = \sqrt{\alpha_1^2 + \alpha_2^2}$.
- Similarly,

$$\frac{\alpha_2}{\alpha_1} = \frac{\sin\phi}{\cos\phi} = \tan\phi,$$

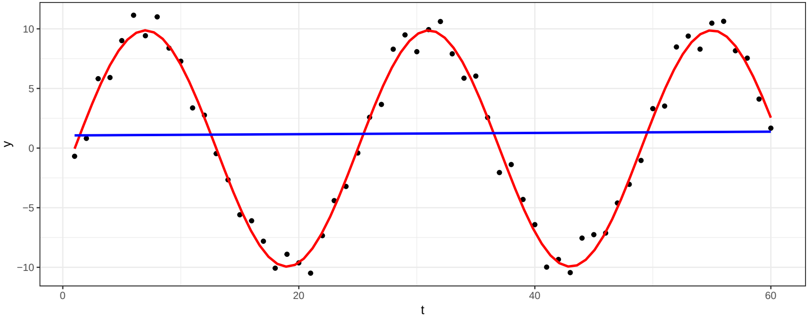
  giving $\phi = \tan^{-1}\left(\frac{\alpha_2}{\alpha_1}\right)$.
- So given $\hat{\alpha}_1, \hat{\alpha}_2, \hat{B}$ we can work back to get $\hat{A}, \hat{B}$ and $\hat{\phi}$.

Meaning of the parameters

- A is the amplitude of the wave, the maximum value the wave takes.
- B is the overall average.
- ϕ is the phase, an offset term, of how the model is shifted from the origin.
- ω is the frequency. This encodes over what time the period repeats.

Linear regression after transformation.

```
sim_plot <- ggplot(sim) + aes(t, y) +
  geom_point() +
  geom_smooth(formula =
                y ~ sin(2*pi*x/24) + cos(2*pi*x/24),
              color = "red", method = "lm", se=FALSE)
sim_plot + geom_smooth(method = "lm",
                       color = "blue", se = FALSE)
```



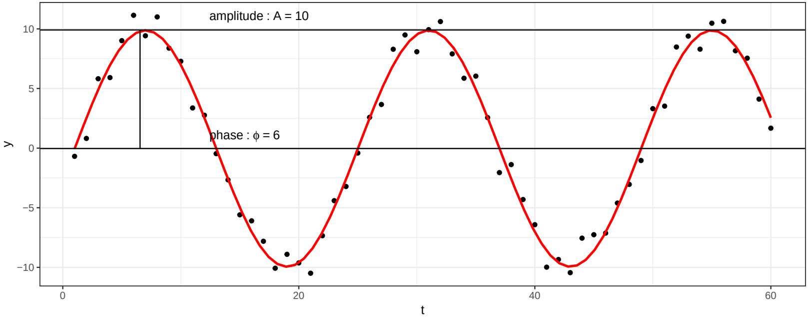
Amplitude, frequency, and phase parameterize sinusoids

```
fit <- sim %>%
  mutate(a1 = sin(2*pi*t/24),
         a2 = cos(2*pi*t/24)) %>%
  lm(y ~ a1 + a2, .) %>%
  coef() %>%
  t() %>%
  as_tibble() %>%
  transmute(B=`(Intercept)`,
            f=atan(a2/a1),
            A=sqrt(a2^2+a1^2))
fit %>% kable(digits = 2)
```

B	f	A
-0.03	-0.26	9.91

Amplitude, frequency, and phase parameterize sinusoids

```
sim_plot + geom_hline(yintercept=c(fit$B,fit$A)) +
 geom_segment(aes(x = 2 * pi - fit$f, y = B,
    xend = 2 * pi - fit$f, yend = B+A), data = fit) +
  annotate("text", 15 * pi*-fit$f, 1, hjust = 0,
    label="phase~':'~phi~'='~6", parse= TRUE)+
  annotate("text", 15 * pi*-fit$f, 11, hjust = 0,
         label="amplitude~':'~A~'='~10", parse= TRUE)
```



References

1. Brockwell, Peter J., Richard A. Davis, and Matthew V. Calder. Introduction to time series and forecasting. Vol. 2. New York: springer, 2002.