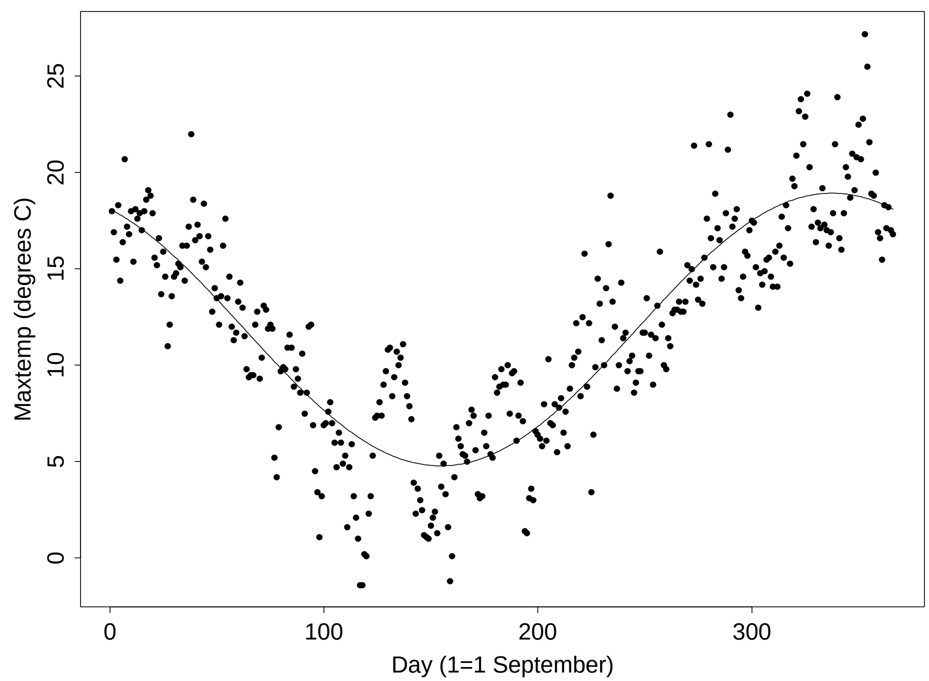
A meteorological time series

- maximum daily temperatures (degrees C) at Bailrigg field-station, September 1995 to August 1996
- note that an unusually cold Christmas 1995 was followed by a mild perod in January-February



Points for discussion

- **what are the main features of the data?**
- **how did I fit the smooth curve to the data?**
- **what features are and are not explained by the fitted curve?**

A harmonic regression model

$$
\begin{aligned}
Y(t) &= \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual} \\
&= \mu + \beta_1 \cos(2\pi t/p) + \beta_2 \sin(2\pi t/p) + \text{residual}
\end{aligned}
$$

- μ = overall mean value (of time series $Y(t)$)
- p = period
- α = amplitude
- ϕ = phase

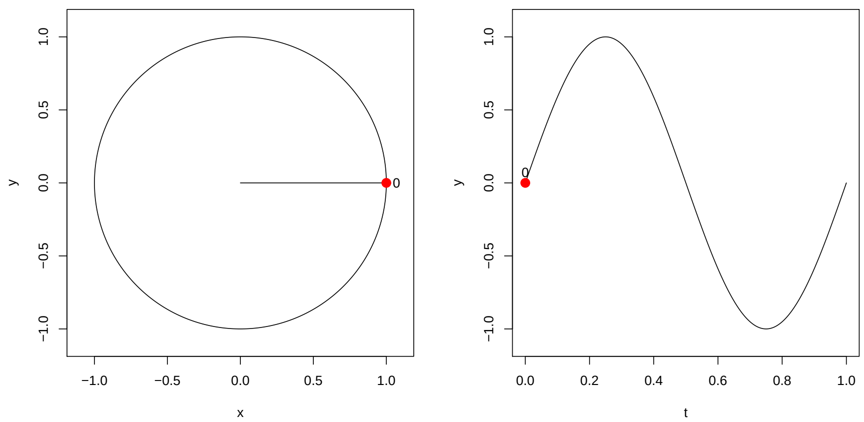
Usually, period is known, but mean, amplitude, phase are unknown.

Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

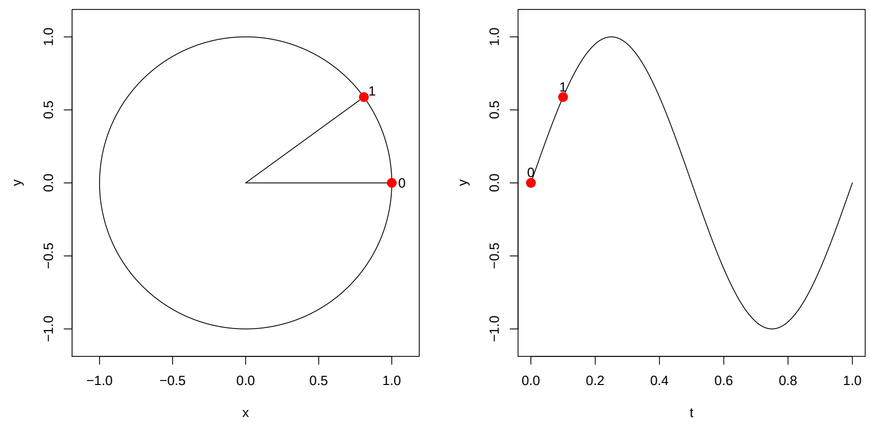


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

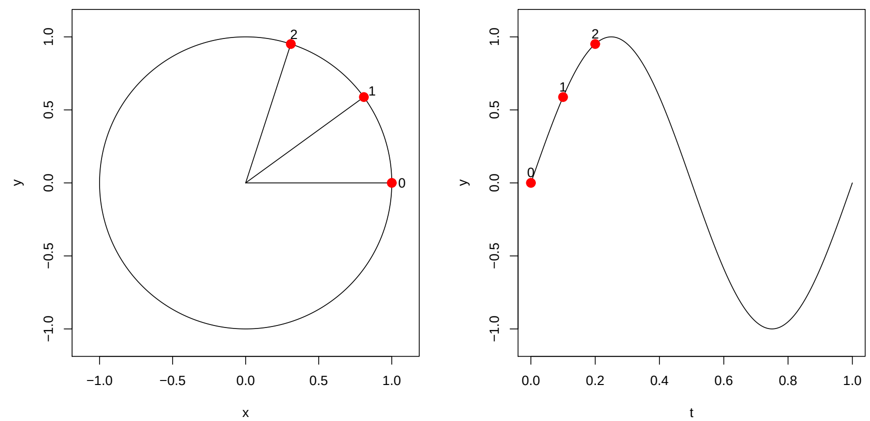


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

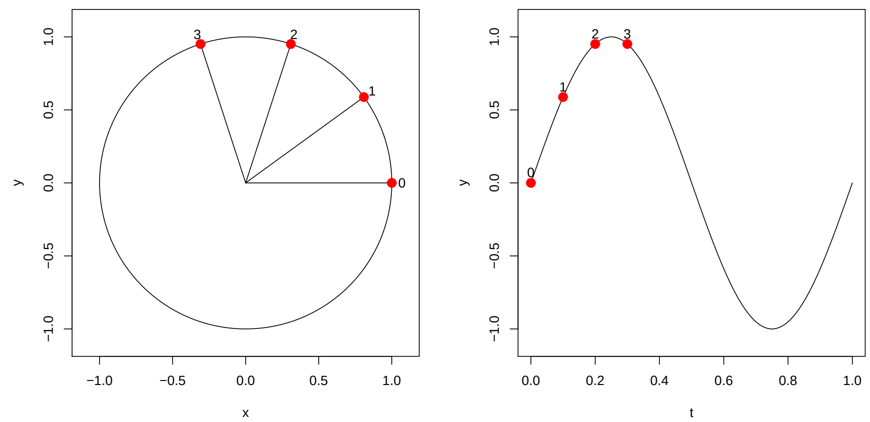


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

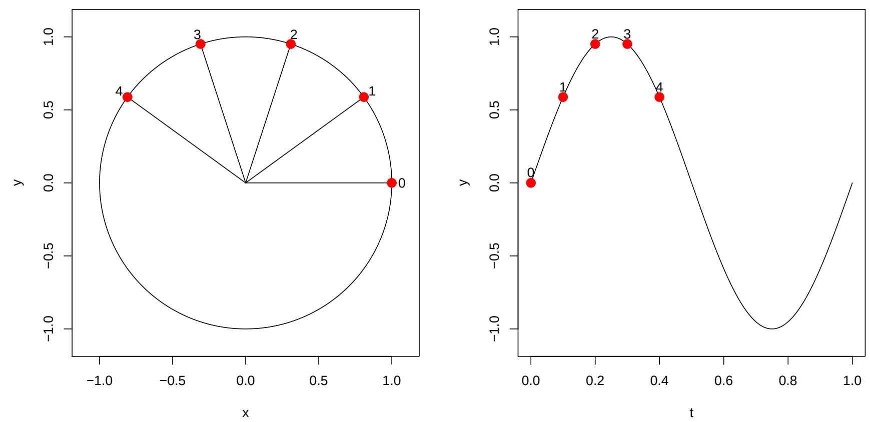


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

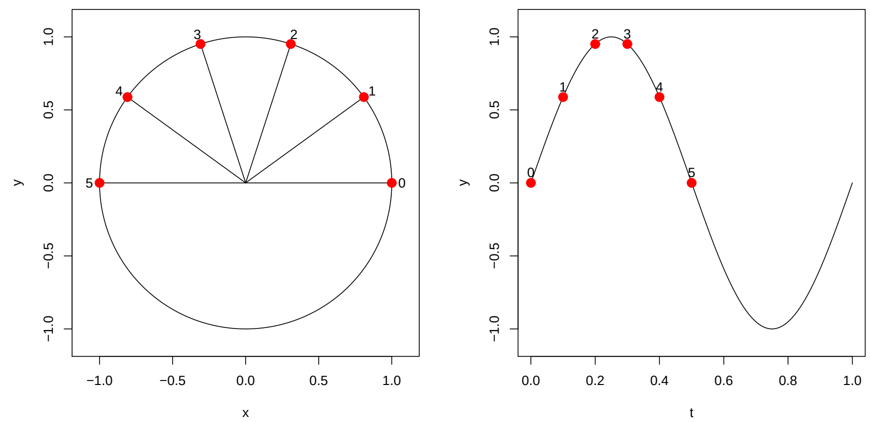


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

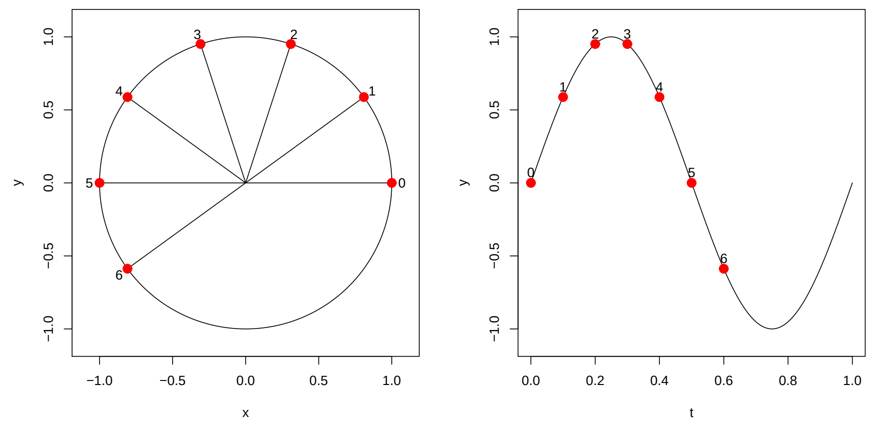


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

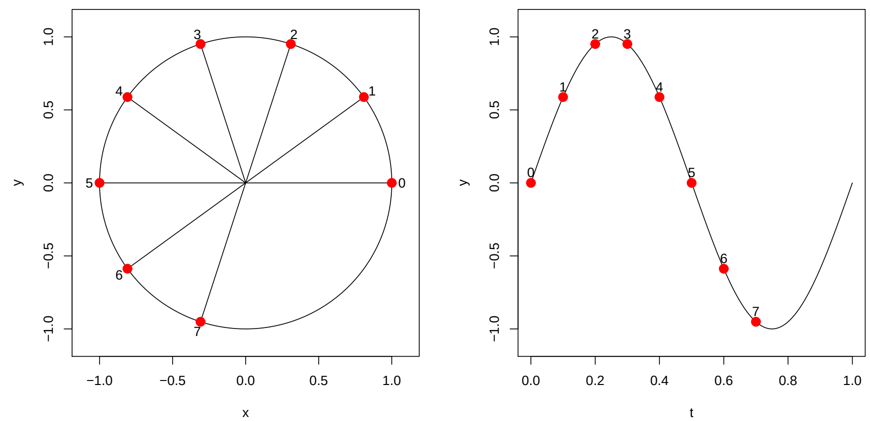


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

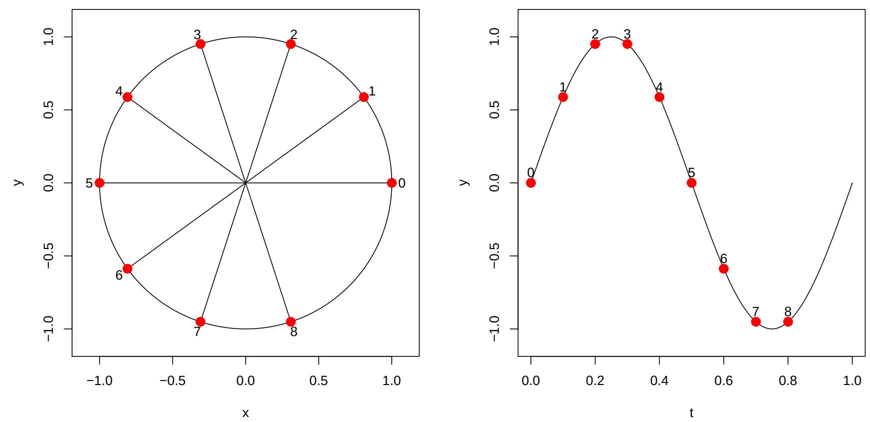


Why does the model work?

Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle

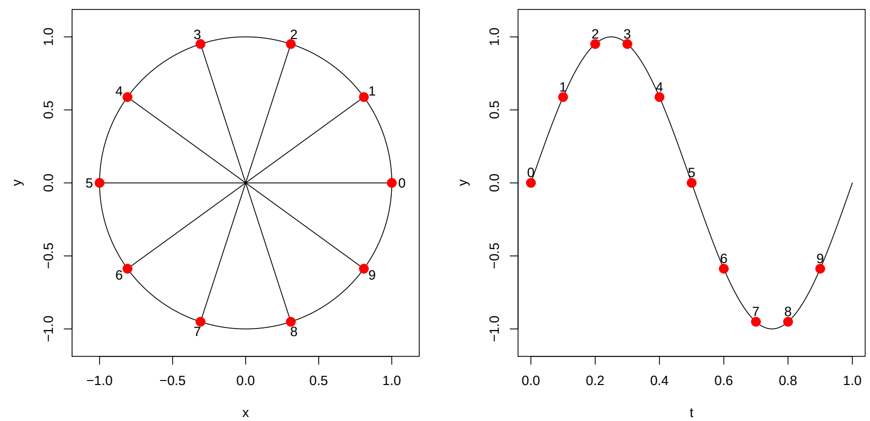


Why does the model work?

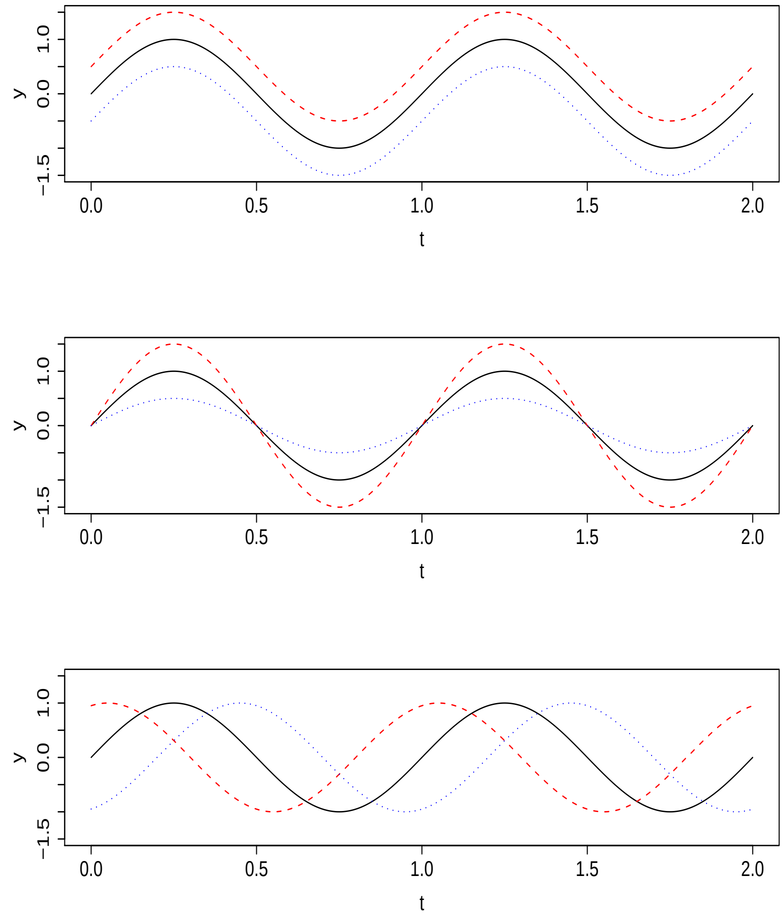
Use the first form of the model,

$$Y(t) = \mu + \alpha \cos(2\pi t/p + \phi) + \text{residual}$$

Now imagine tracking the vertical displacement of a particle moving at constant speed around a circle



Lifting, stretching, shifting



μ = **lifting** α = **stretching** ϕ = **shifting**

Fitting the model

Use the second form of the model,

$$Y(t) = \mu + \beta_1 \cos(2\pi t/p) + \beta_2 \sin(2\pi t/p) + \text{residual}$$

Note that the following quantities are known, i.e. they can be calculated without having to estimate anything

- $x_1(t) = \cos(2\pi t/p)$
- $x_2(t) = \sin(2\pi t/p)$

Re-write the model as

$$Y = \mu + \beta_1 x_1 + \beta_2 x_2$$

After fitting (see next page), amplitude and phase can be recovered using

$$\alpha = \sqrt{\beta_1^2 + \beta_2^2} \qquad \phi = \tan^{-1}(\beta_2/\beta_1)$$

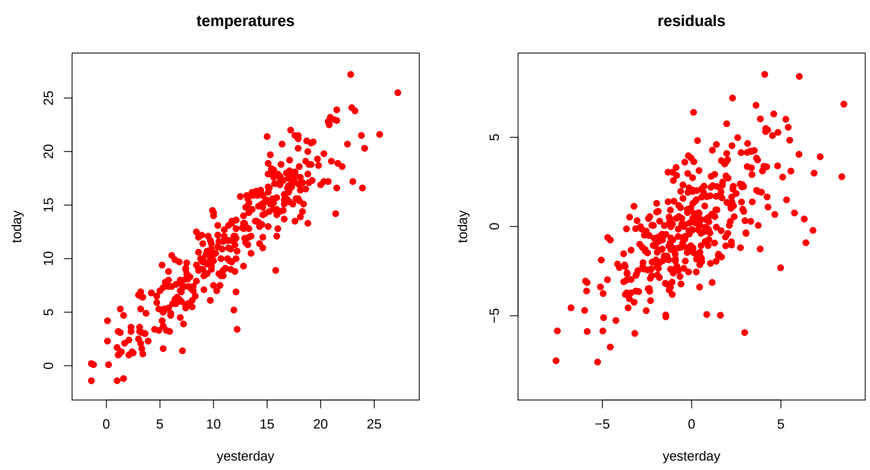
Using the `lm()` function to fit the model

```
data<-read.table("maxtemp.dat")
y<-data[,4]
day<-1:366
x1<-cos(2*pi*day/366)
x2<-sin(2*pi*day/366)
fit<-lm(y~x1+x2)
summary(fit)
```

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-7.5921 -1.8240 -0.1475  1.7140  8.5232

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.8467     0.1441   82.22   <2e-16 ***
x1            6.2508     0.2038   30.68   <2e-16 ***
x2           -3.3177     0.2038  -16.28   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.756 on 363 degrees of freedom
Multiple R-Squared: 0.7687,     Adjusted R-squared: 0.7674
F-statistic: 603.1 on 2 and 363 DF,  p-value: < 2.2e-16
```

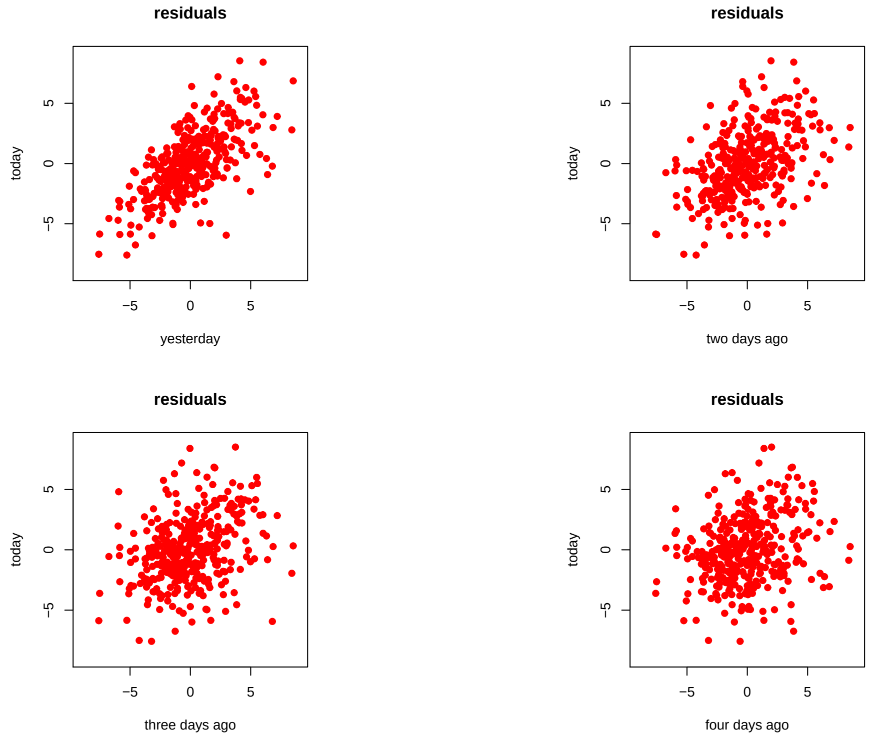
Autocorrelation

- what is the relationship between today's and yesterday's temperature?
- what is the relationship between today's and yesterday's residual?



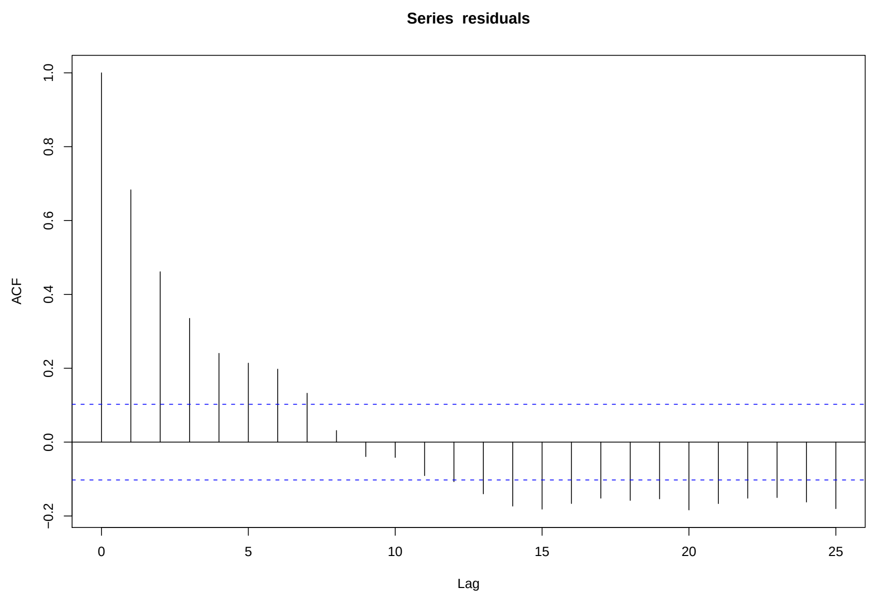
- how and why are the two relationships different?

- **how does the relationship between residuals today and k days ago change as k increases?**



- **lag-k autocorrelation** is the correlation between pairs of values from the same time series k time-units apart

- **correlogram** is a plot of lag-k autocorrelation against k



- dashed lines at $\pm 2/\sqrt{n}$ are pointwise 95% limits for uncorrelated residuals – but overall pattern is more important than individual numerical values

Exercise

Imagine that you have data on the daily maximum temperatures up to today.

- **how would make a forecast of:**
  - **tomorrow's temperature?**
  - **the temperature one month from now?**
- **in what ways are your two answers different, and why?**

Time series models and random effects

$$Y(t) = \mu + \beta_1 \cos(2\pi t/p) + \beta_2 \sin(2\pi t/p) + \text{residual}(t)$$

- $\mu(t) = \mu + \beta_1 \cos(2\pi t/p) + \beta_2 \sin(2\pi t/p)$
- $\text{residual}(t) = S(t) + Z_t$
  - $\text{Cov}\{S(t), S(t-u)\} = \sigma^2 \exp(-u/\phi)$ (random effect)
  - Z_t uncorrelated $\text{N}(0, \tau^2)$ (measurement error)

$$Y(t) = \mu(t) + S(t) + Z_t$$

Using the geoR package to fit the random effects model

You can fool a spatial statistics package to analyse time series data (should you want to) by adding a dummy coordinate

Set-up

```
library(geoR)
maxtemp<-read.table("../datasets/maxtemp_data.txt",header=TRUE)
maxtemp$day<-1:366
maxtemp$zero<-rep(0,366)
maxtemp$xc<-cos(2*pi*maxtemp$day/366)
maxtemp$xs<-sin(2*pi*maxtemp$day/366)
names(maxtemp)
#[1] "year" "m"    "d"    "temp" "day"  "zero" "xc"   "xs"
```

Fit

```
maxtemp<-as.geodata(maxtemp[,c(5,6,4,7,8)],covar.col=4:5)
sigmasq.init<-2.8*2.8; phi.init<-3
mlfit<-likfit(maxtemp,ini.cov.pars=c(sigmasq.init,phi.init,
   nugget=0.5,trend=~xc+xs,cov.model="matern",kappa=0.5)
betahat<-mlfit$beta
betahat.se<-sqrt(diag(mlfit$beta.var))
betahat.corr<-mlfit$beta.var/outer(betahat.se,betahat.se,"*")
round(cbind(betahat,betahat.se,betahat.corr),3)
```

Regression parameters

```
#              betahat betahat.se     V1     V2 V3
#intercept  11.839       0.327  1.000 -0.016  0
#covar1       6.236       0.460 -0.016  1.000  0
#covar2      -3.318       0.465  0.000  0.000  1
```

Covariance parameters

```
mlfit$cov.pars
#[1] 7.503618 2.610402
mlfit$nugget
#[1] 0
```

Compare with results ignoring autocorrelation

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.8467     0.1441   82.22   <2e-16 ***
x1             6.2508     0.2038   30.68   <2e-16 ***
x2            -3.3177     0.2038  -16.28   <2e-16 ***
```