Time Series Regression - Part I

STAT 1321/2320

Kiran Nihlani

Department of Statistics
University of Pittsburgh

Outline

Trend → Trend + seasonal trend.

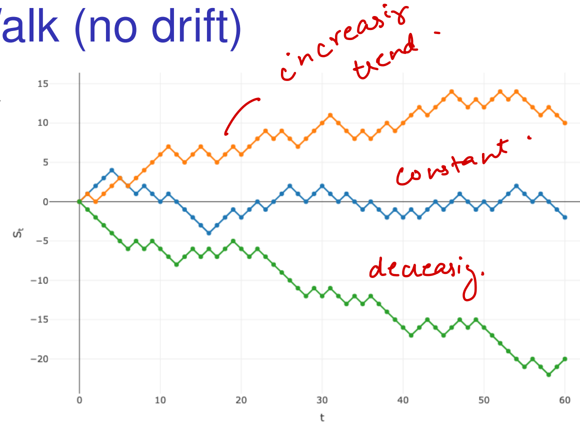
- Time series that exhibit trends over time have a mean function that is some simple function of time.
  - The mean function may or may not be constant.

mean = $f(t)$ → $\hat{\mu}_t$.

- The trend can be deterministic or stochastic.
  - A stochastic trend is “random" and not fundamental to the underlying process.
    - ★ An example of a stochastic trend is a random walk.
  - A deterministic trend is more important and fundamental to the time series process.
    - ★ A linear, quadratic, or even a periodic seasonal trend.

Stochastic Trend - Random Walk (no drift)

- Three realizations of a symmetric binary random walk are presented here.
- The first graph shows an upward trend while the third is decreasing trend.
- We know that a random walk process has constant mean zero.



The upward and downward trend, therefore, is simply a characteristic of that one random realization of the random walk.

Such "trends" could be called stochastic trends, since they are just random.

W_t : white noise.

$$X_t = \sum_{j=1}^{t} W_j$$

random walk.

$X_1 = W_1$

$X_2 = W_1 + W_2$

$\vdots$

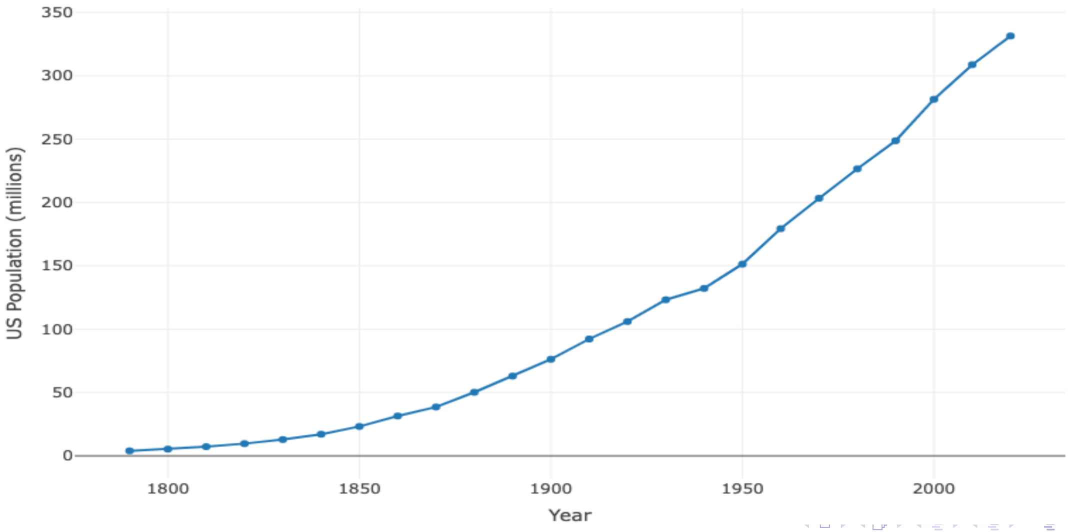
$X_t = W_1 + W_2 + \cdots + W_t$

Binary symm r.w.

$$W_t = \begin{cases} -1 & p = 0.5 \\ 1 & p = 0.5 \end{cases}$$

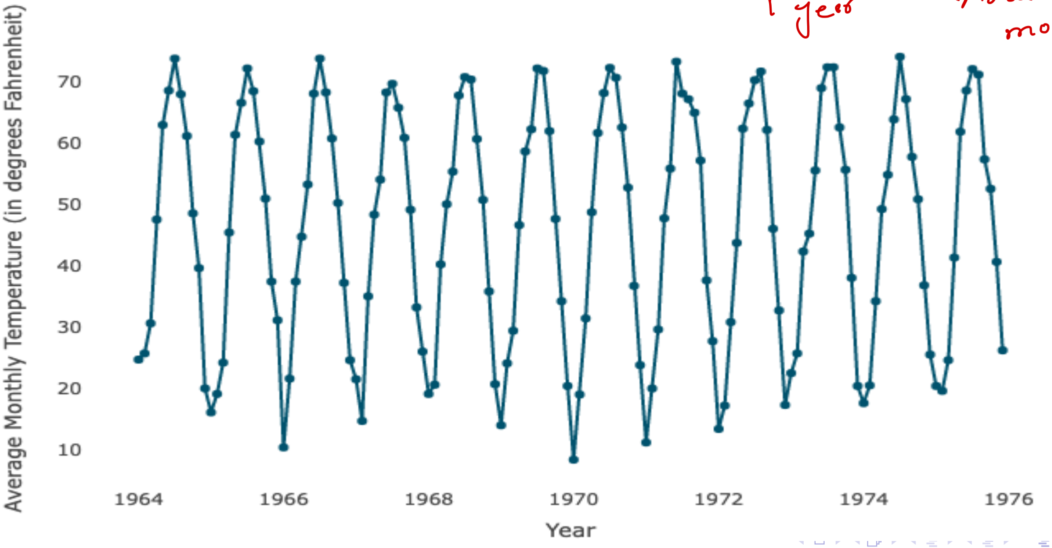
$E(W_t) = 0$, $V(W_t)$ finite.

US Population Census Data



Dubuque, Iowa Temperature Data

12 months
1 year
Additive model.



Deterministic Trend

$X_t = T_t + S_t + I_t$.

$E(I_t) = 0$.

A possible model for both the time series could be

$\hat{\mu_t}$

mean fn. error. $E(X_t) = \underbrace{E(T_t + S_t)}_{} = \mu_t$.

$$X_t = \mu_t + \epsilon_t$$

- For the US population data, a quadratic trend may be feasible: $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$
- For the Dubuque temperature data, μ_t would be periodic with period 12

$$\mu_t = \mu_{t-12}$$

We might assume that ϵ_t, the unobserved variation around μ_t, has zero mean for all t so that μ_t is the mean function for the observed series X_t

Linear Trend

X_t: response variable.

explanatory variable.

- A linear trend is expressed as:

$$\underbrace{E(X_t)}_{}\ \mu_t = \beta_0 + \beta_1 t$$

- The least squares method chooses the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the least squares criterion:

obs. expl pred. OLS.

$$\text{Loss } f^n = Q(\beta_0, \beta_1) = \sum_{t=1}^{n} [X_t - (\beta_0 + \beta_1 t)]^2$$

- The expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$ can be found using calculus.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{X}_t - \bar{t}$$

$$\hat{\beta}_1 = \frac{r s_y}{s_x} = \frac{r\, SD(X_t)}{SD(t)}$$

Quadratic Trend

$X_t = \mu_t + \varepsilon_t$: linear reg.

- A linear model with quadratic trend is expressed as:

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

- The least squares method chooses the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ that minimize the least squares criterion: obs - pred

$$Q(\beta_0, \beta_1, \beta_2) = \sum_{t=1}^{n} [X_t - \underbrace{(\beta_0 + \beta_1 t + \beta_2 t^2)}]^2$$

- Before fitting a linear (or quadratic) model, it is important to ensure that this trend truly represents the deterministic nature of the time series process.

Global vs Local Trend

- The deterministic trend equation of these models is based on the entire dataset and hence sometimes called a **global** trend model.
- This may be unrealistic in cases where the series fluctuates a lot and/or has outliers.
- A **local** trend model (like a piece-wise linear model) may be preferred in cases like this.

Example- US Population

- Fit a quadratic model to the US population data. The R output is given below:

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
Intercept	8.05082 $\hat{\beta}_0$	2.01809	3.989	0.000667 ***
time	-2.52389 $\hat{\beta}_1$	0.37195	-6.786	1.04e-06 ***
time^2	0.67082 $\hat{\beta}_2$	0.01444	46.443	$<$ 2e-16 ***

$H_0: \beta_0 = 0$

$H_0: \beta_1 = 0$

$H_0: \beta_2 = 0$

- The time variable was coded as $1, 2, 3, \cdots$. We could also have used $t = 1790, 1800, \cdots$

$R^2 \rightarrow 1.$

- The R-squared for the model is 0.9992.
- How do you interpret the regression coefficients?

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t.$$

→ create time vector.

$t = c(1, 2, 3, \ldots, n)$

$t = c(1790, 1800, 1810, \ldots, 2020)$

→ $lm(X_t \sim t + t^2)$

$t \to 0 \to 1$

avg chang. in y is β_1

"y_t" $= \beta_0 + \beta_1 t$. → β_1

β_0: avg "y_t" at $t = 0$.

$X_t = \beta_0 + \beta_1 t + \beta_2 t^2$. →

β_0: avg. X_t at $t = 0$.

$X_t = \beta_0 + \beta_1 t$.

β_1 is the avg. change in X_t for a unit change in t.

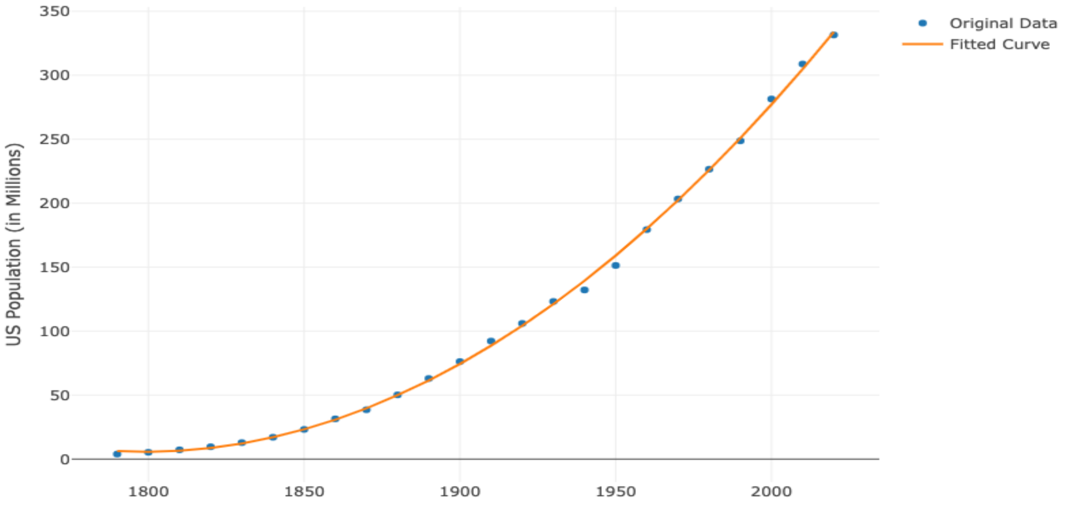
$$\frac{dX_t}{dt} = \beta_1$$

$X_t = \beta_0 + \beta_1 t + \beta_2 t^2$.

$$\frac{dX_t}{dt} = \beta_1 + 2\beta_2 t$$

for a unit change in "t", the avg. change in X_t is $\underline{\beta_1 + 2\beta_2 t}$

Example- US Population



Seasonal/Cyclical Trends in Data

quarter ·

- The **seasonal means** approach represents the mean function with a different parameter for each level.
- For example, suppose each measured time is a different quarter, and we have observed data over a period of several years.
- The seasonal means model might specify a different mean response for each of the 4 quarters.

$$\mu_t = \begin{cases} \beta_1 & \text{for } t = 1, 5, 9, \ldots \\ \beta_2 & \text{for } t = 2, 6, 10, \ldots \\ \beta_3 & \text{for } t = 3, 7, 11, \ldots \\ \beta_4 & \text{for } t = 4, 8, 12, \ldots \end{cases}$$

Seasonal/Cyclical Trends in Data

- This is similar to an ANOVA model in which the parameters are the mean response values for each factor level.
- The model does not contain an intercept, and that fact needs to be specified in the fitting software.

seasonality.

quarterly setup.

$$X_t = \overbrace{\beta_1 d_{1,t} + \beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t}} + \epsilon_t \quad \beta_5 t + \beta_6 t^2$$

where $d_{i,t}$ is a dummy/indicator variable.

- An alternative formulation does include an intercept and omits one of the β's in the previous model (the interpretations would change of course!).
- The model may accommodate a linear/polynomial trend as well.

Example - Australian Beer Production Data

- We want to forecast the value of future beer production using a regression model with a linear trend and quarterly dummy variables

seasonality. linear trend.

$$X_t = \underbrace{\beta_1 d_{1,t} + \beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t}}_{\text{seasonality}} + \underbrace{\beta_5 t}_{\text{linear trend}} + \epsilon_t$$

- $d_{i,t}$ is a dummy variable that takes the value 1 if an observation falls in quarter i, and 0 otherwise.
- We use a subset of the Ausbeer data from 1957 Q1 to 1973 Q4.
- The time variable was coded as $1, 2, 3, \ldots$.

"ts" object.

season( )

↳ periodic.

Example - Australian Beer Production Data

		Estimate	Std. Error	t value	Pr(>|t|)
β_1	1Q	236.14614	5.11244	46.19	<2e-16 ***
β_2	2Q	187.79228	5.17727	36.27	<2e-16 ***
β_3	3Q	203.96783	5.24320	38.90	<2e-16 ***
β_4	4Q	288.73162	5.31019	54.37	<2e-16 ***
linear trend ← β_5	t	3.05974	0.09978	30.66	<2e-16 ***

- The model shows an average upward trend of 3.0597 mega liters per quarter.
- The average production for Q1 is 236.146 mega litres and so on.

  after adjusting for trend

Example - Australian Beer Production Data

- An alternative formulation of this model can be

$$X_t = \overbrace{\beta_0}^{Q1} + \beta_1 t + \overbrace{\beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t}}^{\text{3 dummy variables}} + \epsilon_t$$

- We fit a model with an intercept term so the software omits the Q1 coefficient in this case.

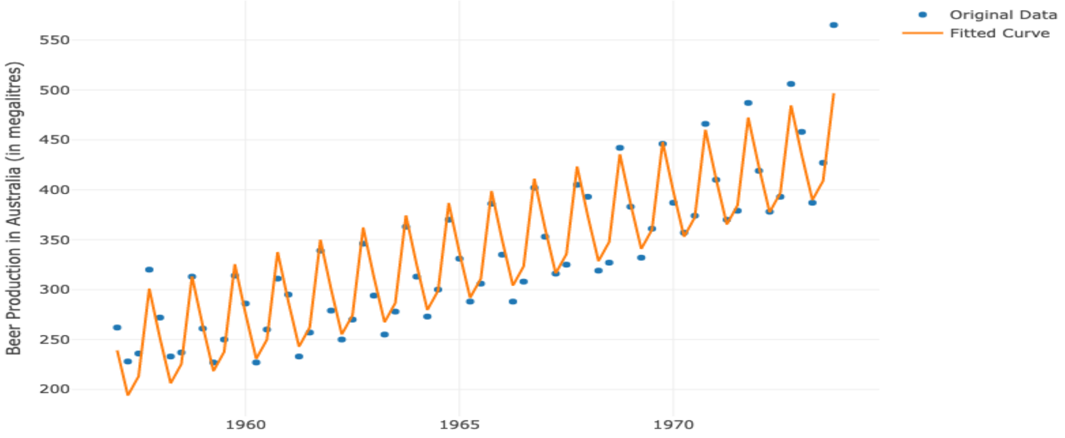
		Estimate	Std. Error	t value	Pr(>|t|)
	(Intercept)	236.14614	5.11244	46.191	< 2e-16 ***
	t	3.05974	0.09978	30.663	< 2e-16 ***
Q2 - Q1	2Q	-48.35386	5.53152	-8.742	1.82e-12 ***
Q3 - Q1	3Q	-32.17831	5.53422	-5.814	2.19e-07 ***
Q4 - Q1	4Q	52.58548	5.53872	9.494	9.12e-14 ***

Example - Australian Beer Production Data

- There is an average upward trend of 3.05974 megalitres per quarter. every year.
- Now the Q2 coefficient is interpreted as the difference between Q2 and Q1 average production, the Q3 coefficient is the difference between Q3 and Q1 average production, and so forth.
- On average,
  - the second quarter has production of 48.35 megalitres lower than the first quarter
  - the third quarter has production of 32.18 megalitres lower than the first quarter
  - the fourth quarter has production of 52.59 megalitres higher than the first quarter.

Example - Australian Beer Production Data

R-squared = 0.998



Harmonic Regression for Cosine Trends

- The *seasonal means* model makes no assumption about the shape of the mean response function over time.
- A more specific model might assume that the mean response varies over time in some regular manner.
- For example, a model for temperature data might assume mean temperatures across time rise and fall in a periodic pattern, such as:

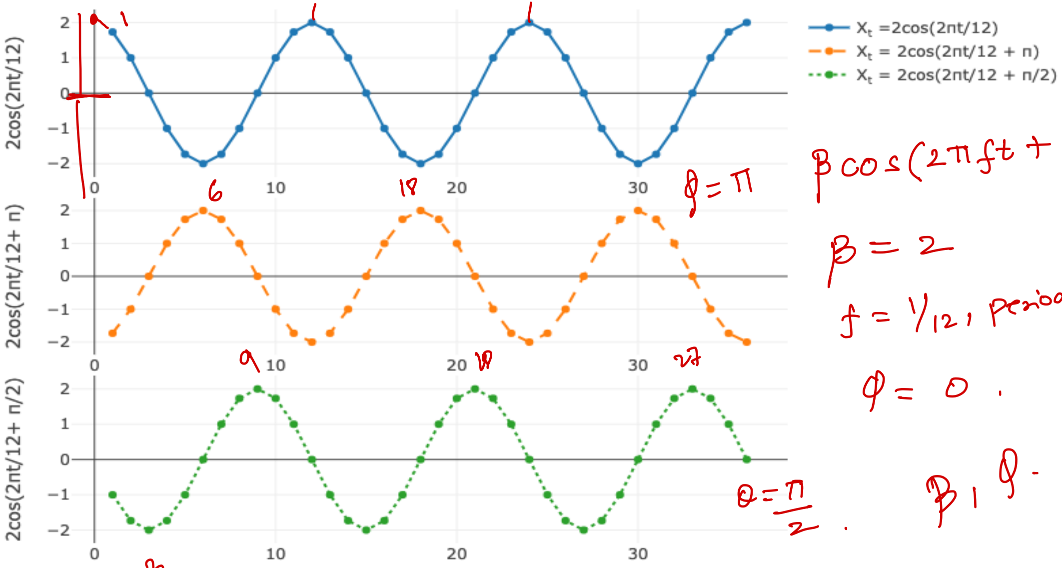
$$\mu_t = \beta \underline{\underline{\cos(2\pi f t + \phi)}}$$

where β is amplitude, f the frequency, and ϕ the phase.

Parameters

- The amplitude β is the height of the cosine curve from its midpoint to its top.
  - As *t* varies, the curve oscillates between a maximum of β and a minimum of $-\beta$.
- The frequency *f* measures how often the curve's pattern repeats itself.
  - *f* is the reciprocal of the period.
  - If monthly data is recorded as $t = 1, 2, \cdots 12, \cdots$, the period = 12 and $f = 1/12$.
  - If the data are monthly but time is measured in years, e.g., $t = 2016, 2016.0833, 2016.1667 \cdots$, then the period is 1 and *f* would be 1 in this case.
- The phase, ϕ, decides the origin and the max/min point.

$X_t = \beta \cos(2\pi f t + \phi)$



$\phi = \pi$

$\beta \cos(2\pi f t + \phi)$

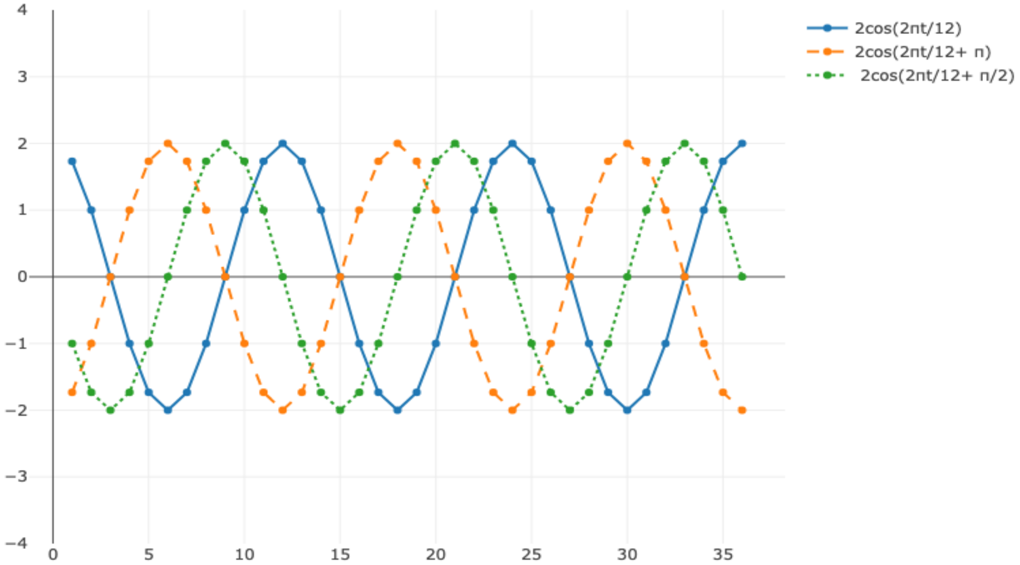
$\beta = 2$

$f = 1/12$, period $= 12$

$\phi = 0$.

$\theta = \frac{\pi}{2}$.

β, ϕ.


2cos(2πt/12)
2cos(2πt/12+ π)
2cos(2πt/12+ π/2)
4
3
2
1
0
−1
−2
−3
−4
0
5
10
15
20
25
30
35

Harmonic Regression

- The mean function set as $\mu_t = \beta \cos(2\pi f t + \phi)$ is inconvenient for estimation because the parameters β and ϕ do not enter the expression linearly (f is usually known or can be easily estimated).
- To fit the model, it is useful to consider a transformation of the mean response function:

$$\mu_t = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

where

$$\beta = \sqrt{\beta_1^2 + \beta_2^2} \text{ and } \phi = \mathit{arctan}(-\beta_2/\beta_1)$$

tan^{-1}( )

atan

atan2

and, conversely,

$$\beta_1 = \beta \cos(\phi), \quad \beta_2 = \beta \sin(\phi)$$

Harmonic Regression

β, ϕ

non-linear

- To estimate the parameters β_1 and β_2 with regression techniques, we simply use $\cos(2\pi ft)$ and $\sin(2\pi ft)$ as regressors or predictor variables.
- The simplest such model for the trend would be expressed as $\beta\cos(2\pi ft+\phi)$

$$\mu_t = \beta_0 + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft)$$

Here the constant term, β_0, can be meaningfully thought of as a cosine with frequency zero.

$\beta_1 = \beta\cos(\phi)$

$\beta_2 = \beta\sin(\phi)$

$\beta = \sqrt{\beta_1^2 + \beta_2^2}$

Example - Dubuque Temperature

Temp. data

- Fitting the following model to the average monthly demperature data for Dubuque, Iowa

mean temp. over time.

$$\mu_t = \boxed{\beta_0} + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft)$$

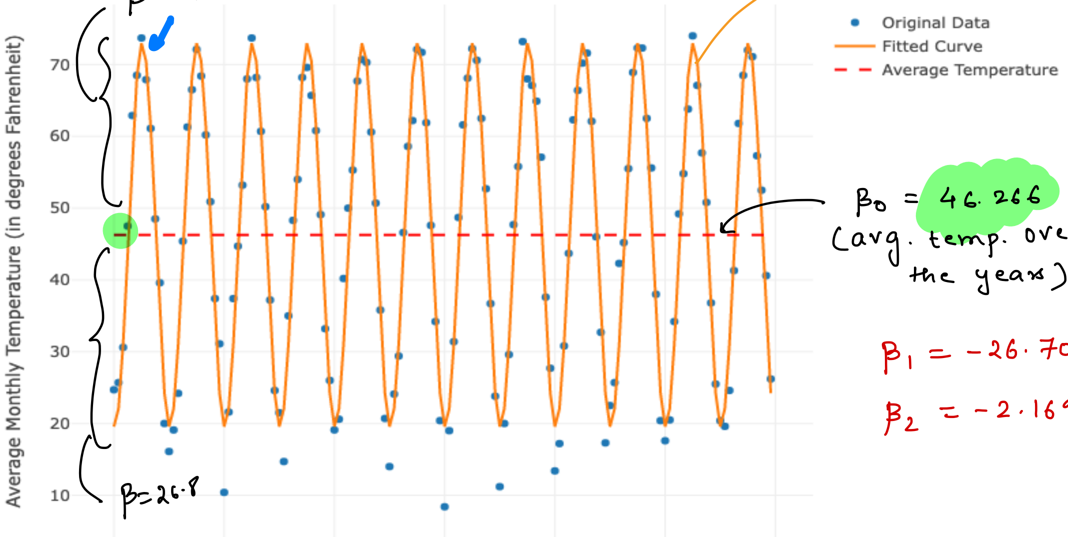
	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	46.2660	0.3088	149.816	< 2e-16 ***
$\cos(2\pi t)$	-26.7079 β_1	0.4367	-61.154	< 2e-16 ***
$sin(2\pi t)$	-2.1697 β_2	0.4367	-4.968	1.93e-06 ***

- $\beta_0 = 46.266$ is the average temperature (baseline) over time.
- Amplitude, $\beta = \sqrt{\beta_1^2 + \beta_2^2} = \sqrt{-26.7079^2 + -2.1697^2} \approx 26.8.$

Example - Dubuque Temperature

$\beta_0 + \beta_1 \cos(2\pi t) + \beta_2 \sin(2\pi t)$

$\beta = 26.8$



$\beta = 26.8$

$\beta_0 = 46.266$ (avg. temp. over the years)

$\beta_1 = -26.7079$

$\beta_2 = -2.1697$

$\beta_1 = \beta \cos(\phi)$ shift phase.

amplitude.

→ cos maxima occurs at $0, 2\pi, \ldots$

→ maxima will occur at $t_0, t_{12}, t_{24} \ldots$

→ we need maxima around $\underbrace{t_6, t_7}_{\text{June, July}}$

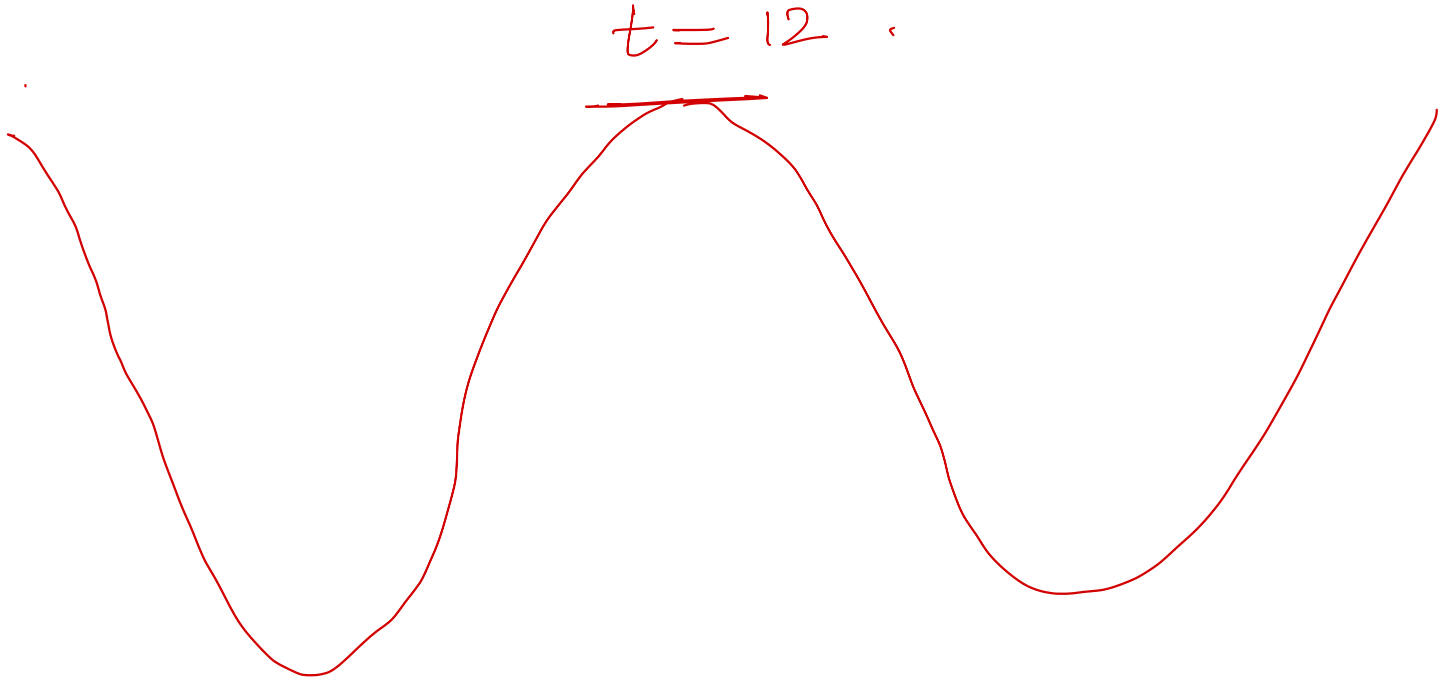
phase shift of π

$\beta = 26.8$

$\beta_1 = -26.7079$

$$\frac{\beta_1}{\beta} \approx -1 = \cos(\phi)$$

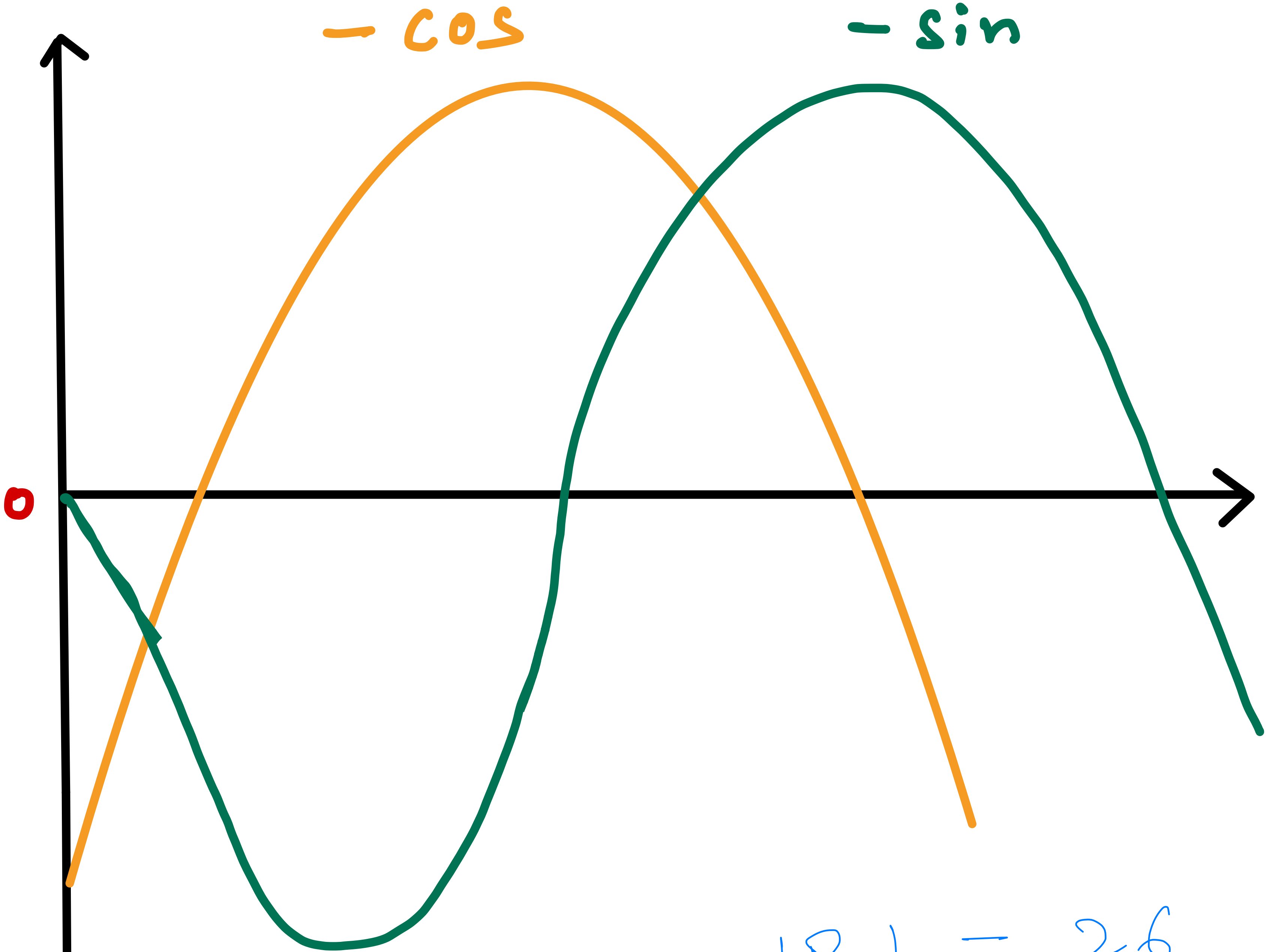
$$\phi = \pi$$



maximum around June

$\pi_{t=6,7}$

$\delta \to \beta_1, \beta_2$

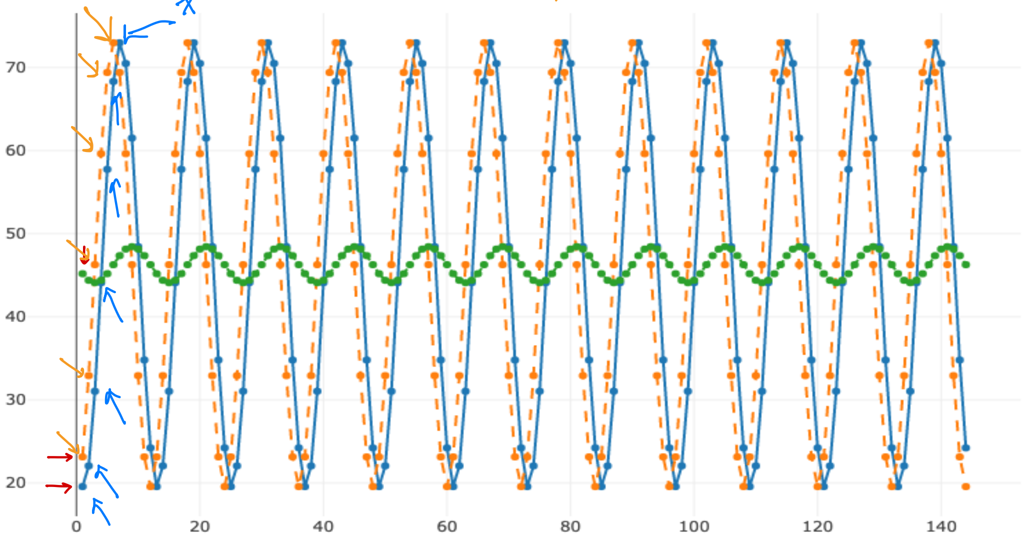


$|\beta_1| = 26$

$|\beta_2| \approx 2 \ldots$

Harmonic Regression - Dubuque, Iowa

- $\beta_0 + \beta_2 \sin(2\pi t)$
- $\beta_0 + \beta_1 \cos(2\pi t)$



Linear Model - Residual Analysis

white noise
" stationary "

We are assuming a linear model which means we are implicitly making some assumptions about the variables.

- First, we assume that the model is a reasonable approximation to reality.
- Second, we make (quite) a few assumptions about the errors $\epsilon_1, \epsilon_2, \epsilon_3, \ldots$
  - They have mean zero; otherwise the forecasts will be systematically biased.
  - They are not autocorrelated; otherwise the forecasts will be inefficient, as there is more information in the data that can be exploited.
  - They are unrelated to the predictor variables; otherwise there would be more information that should be included in the systematic part of the model.
  - They may be assumed to be normally distributed for ease of creating prediction intervals/