THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

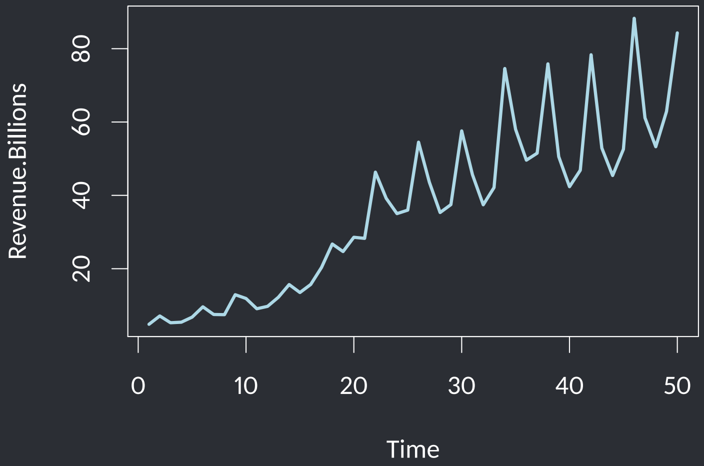
Model building: time and seasonality

Lecture 15

STA 371G

Let's try to forecast Apple's quarterly revenue, in billions of dollars:

```
plot(Revenue.Billions ~ Time, data=apple,
  type="l", col="lightblue", lwd=3)
```



What do we see here?

- ✓ A trend component: revenue is increasing over the long run.

What do we see here?

- ✓ A trend component: revenue is increasing over the long run.
- ✓ A seasonal component: revenue is higher in some quarters than others.

What do we see here?

- ✓ A trend component: revenue is increasing over the long run.
- ✓ A seasonal component: revenue is higher in some quarters than others.
- ✗ No cyclic component: things just seem to be going up over time. (A cyclic pattern would consist of unpredictable short-term trends, like the value of the Dow Jones index over time.)

What do we see here?

- ✓ A trend component: revenue is increasing over the long run.
- ✓ A seasonal component: revenue is higher in some quarters than others.
- ✗ No cyclic component: things just seem to be going up over time. (A cyclic pattern would consist of unpredictable short-term trends, like the value of the Dow Jones index over time.)
- ✓ An irregular component: there is definitely quarter-by-quarter variation that is not accounted for by the other components. (This is the part that can't be modeled!)

Based on our analysis, it seems like a reasonable model would look something like this:

$$\text{Revenue} = \text{Trend} + \text{Seasonality} + \text{Error}$$

This looks a lot like a regression model!

Using regression to model time series

When we use regression to model time series, we almost always violate the independence assumption!

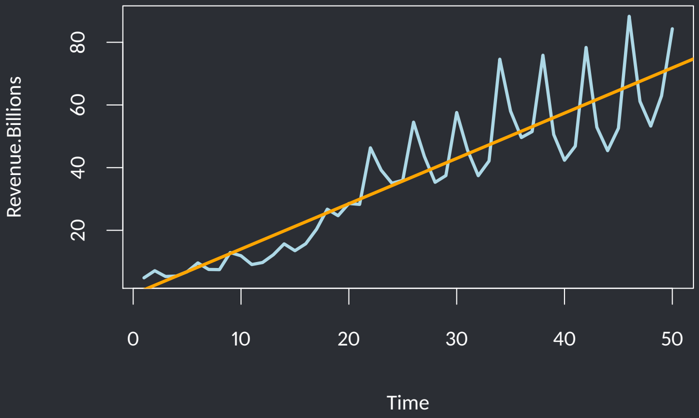
Using regression to model time series

When we use regression to model time series, we almost always violate the independence assumption!

That's OK as long as we don't want to any inference (i.e., use the p-values or construct confidence intervals). Usually with time series our main goal is forecasting.

Take 1: Model with a trend component

```
trend.model <- lm(Revenue.Billions ~ Time, data=apple)
plot(Revenue.Billions ~ Time, data=apple,
  type="l", col="lightblue", lwd=3)
abline(trend.model, col="orange", lwd=3)
```



Take 1: Making predictions

The prediction equation is:

$$\widehat{\text{Revenue}} = -0.42 + 1.45 \cdot \text{Time}$$

To extrapolate the model out into the future, we just have to figure out what the value of the Time variable is for the time period we want to forecast.

Take 1: Making predictions

The prediction equation is:

$$\widehat{\text{Revenue}} = -0.42 + 1.45 \cdot \text{Time}$$

To extrapolate the model out into the future, we just have to figure out what the value of the Time variable is for the time period we want to forecast. For example, the last time period is Q4 2018, which is Time = 50, so Q1 2019 is Time = 51:

$$\widehat{\text{Q2 2019 revenue}} = -0.42 + 1.45 \cdot 51$$

Take 2: Model with trend and seasonal components

What's wrong with this?

```
seasonal.model <- lm(Revenue.Billions ~ Time + Quarter,
  data=apple)
```

Take 2: Model with trend and seasonal components

What’s wrong with this?

```
seasonal.model <- lm(Revenue.Billions ~ Time + Quarter,
  data=apple)
```

It treats Quarter as a quantitative variable, which implies a linear relationship between Quarter and Revenue, which we can see is not true (revenue is lowest in Q2, not Q1):

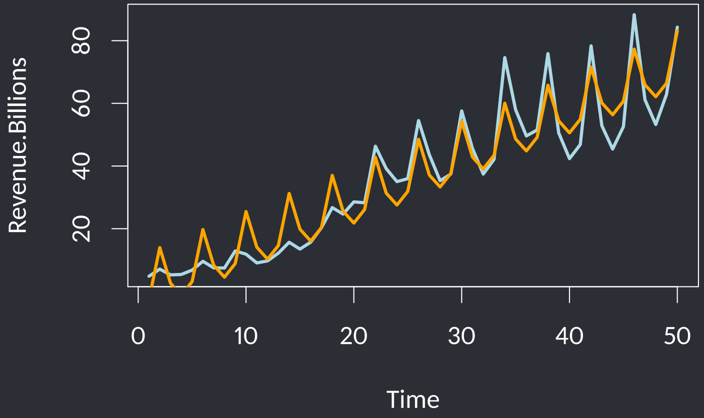
```
tapply(apple$Revenue.Billions, apple$Quarter, mean)

    1     2     3     4 
34.26 30.44 31.90 48.53 
```

We need to tell R that Quarter should be treated as a categorical variable (what R calls a “factor”):

```
apple$QuarterCat <- as.factor(apple$Quarter)
seasonal.model <- lm(Revenue.Billions ~ Time +
  QuarterCat, data=apple)
```

We need to tell R that Quarter should be treated as a categorical variable (what R calls a “factor”):

```
apple$QuarterCat <- as.factor(apple$Quarter)
seasonal.model <- lm(Revenue.Billions ~ Time +
  QuarterCat, data=apple)
```



Take 3: a multiplicative model

Our predictions seem to be overestimating the seasonal fluctuations early on, and underestimating them in more recent years, because our linear model assumes that the effect of each season is an additive constant each year.

Take 3: a multiplicative model

Our predictions seem to be overestimating the seasonal fluctuations early on, and underestimating them in more recent years, because our linear model assumes that the effect of each season is an additive constant each year.

A multiplicative model, where we estimate Y as a function of the product of trend, seasonality, and irregular (error) components, could help:

$$\text{Revenue} = (\text{Trend})(\text{Seasonality})(\text{Error})$$

Take 3: a multiplicative model

Our predictions seem to be overestimating the seasonal fluctuations early on, and underestimating them in more recent years, because our linear model assumes that the effect of each season is an additive constant each year.

A multiplicative model, where we estimate Y as a function of the product of trend, seasonality, and irregular (error) components, could help:

$$\text{Revenue} = (\text{Trend})(\text{Seasonality})(\text{Error})$$

How do we model that with regression?

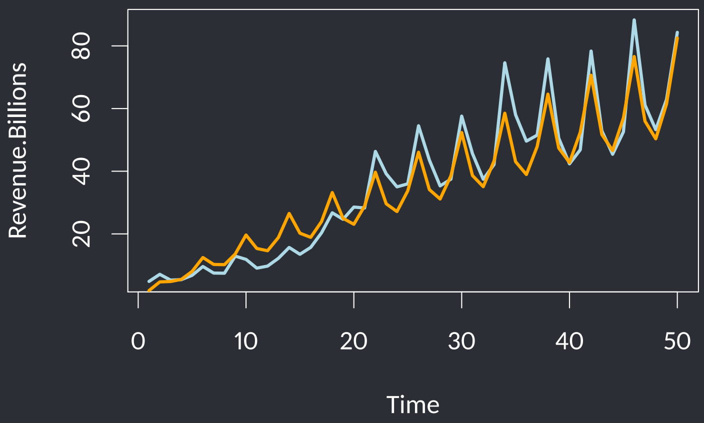
Take 3: a multiplicative model

Take the log of both sides:

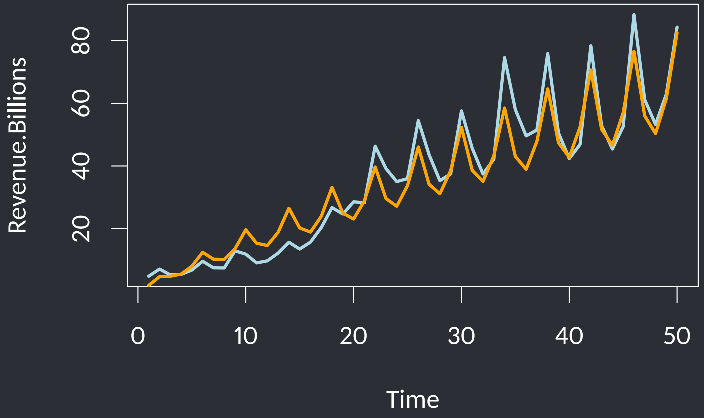
$$\begin{aligned} \text{Revenue} &= (\text{Trend})(\text{Seasonality})(\text{Error}) \\ \log \text{Revenue} &= \log\left((\text{Trend})(\text{Seasonality})(\text{Error})\right) \\ \log \text{Revenue} &= \log \text{Trend} + \log \text{Seasonality} + \log \text{Error} \end{aligned}$$

```r
mult.model <- lm(log(Revenue.Billions) ~ log(Time) +
  QuarterCat, data=apple)
```

```r
mult.model <- lm(log(Revenue.Billions) ~ log(Time) +
  QuarterCat, data=apple)
```



```r
mult.model <- lm(log(Revenue.Billions) ~ log(Time) +
  QuarterCat, data=apple)
```



In Q4 2011, the revenue jumped to $46.33B, by far the highest quarterly revenue ever! Our model is not accounting for this jump—what happened?

SECTIONS HOME SEARCH The New York Times SUBSCRIBE NOW LOG IN

Elizabeth Warren Proposes Breaking Up Tech Giants Like Amazon and Facebook

The Man Deciding Facebook's Fate

THE NEW NEW WORLD
Mark Zuckerberg Wants Facebook to Emulate WeChat. Can It?

BITS
The Week in Tech: Facebook's Privacy Pivot (Business Model Not...

India F Blindn A.I.

TECHNOLOGY

Jobs Steps Down at Apple, Saying He Can't Meet Duties

By DAVID STREITFELD AUG. 24, 2011

SAN FRANCISCO — Steven P. Jobs, whose insistent vision that he knew what consumers wanted made Apple one of the world's most valuable and influential companies, is stepping down as chief executive, the company announced late Wednesday.

RELATED COVERAGE

While You Were Out: Apple's Years With and

Take 4: incorporating Tim Cook

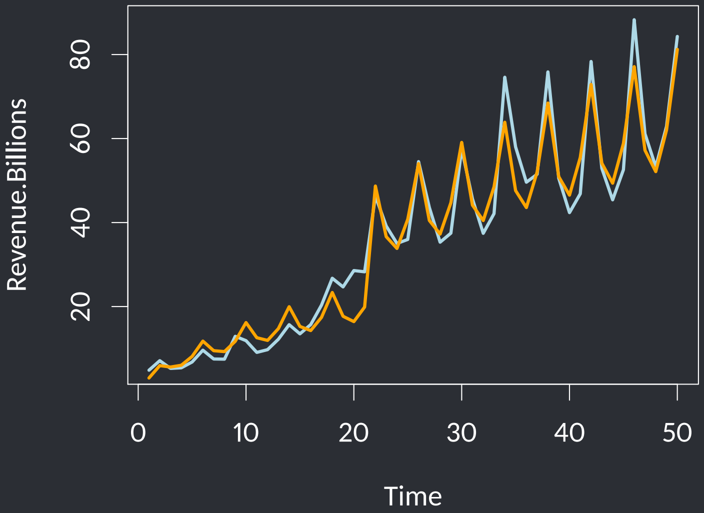
Let's define a dummy variable that is 1 when Tim Cook is CEO for the full quarter (when Time $\geq$ 22; i.e., starting in Q4 2011) and 0 otherwise:

```
apple$TimCookEra <- ifelse(apple$Time >= 22, 1, 0)
```

Then, let's add this as an additional predictor variable to the model:

```
mult.model2 <- lm(log(Revenue.Billions) ~ log(Time) +
  QuarterCat + TimCookEra, data=apple)
```

Take 4: incorporating Tim Cook



Which model is best?

We can use R^2 to compare models, as usual:

Model	R^2
Trend only (additive)	0.8212
Trend + Seasonal (additive)	0.9132
Trend + Seasonal (multiplicative)	0.9028
Trend + Seasonal + Tim Cook (multiplicative)	0.9524

Which model is best?

We can use R^2 to compare models, as usual:

Model	R^2
Trend only (additive)	0.8212
Trend + Seasonal (additive)	0.9132
Trend + Seasonal (multiplicative)	0.9028
Trend + Seasonal + Tim Cook (multiplicative)	0.9524

Another approach is to use the average absolute prediction error (or average percent prediction error) when predicting revenue at time $t+1$ using only the data from time $1, 2, \ldots, t$.

What about modeling cyclic components?

- When there is a cyclic component (e.g., due to business cycles, like the ups and downs of the stock market), neither a trend nor seasonal model will be appropriate.
- To model cyclic time series you can use autoregression, where we predict what happens in time t using what happens in time $t-1, t-2, \ldots$. (We won't cover this in 371!)