Week 1: Descriptive Analytics

- An Operational Decision Problem

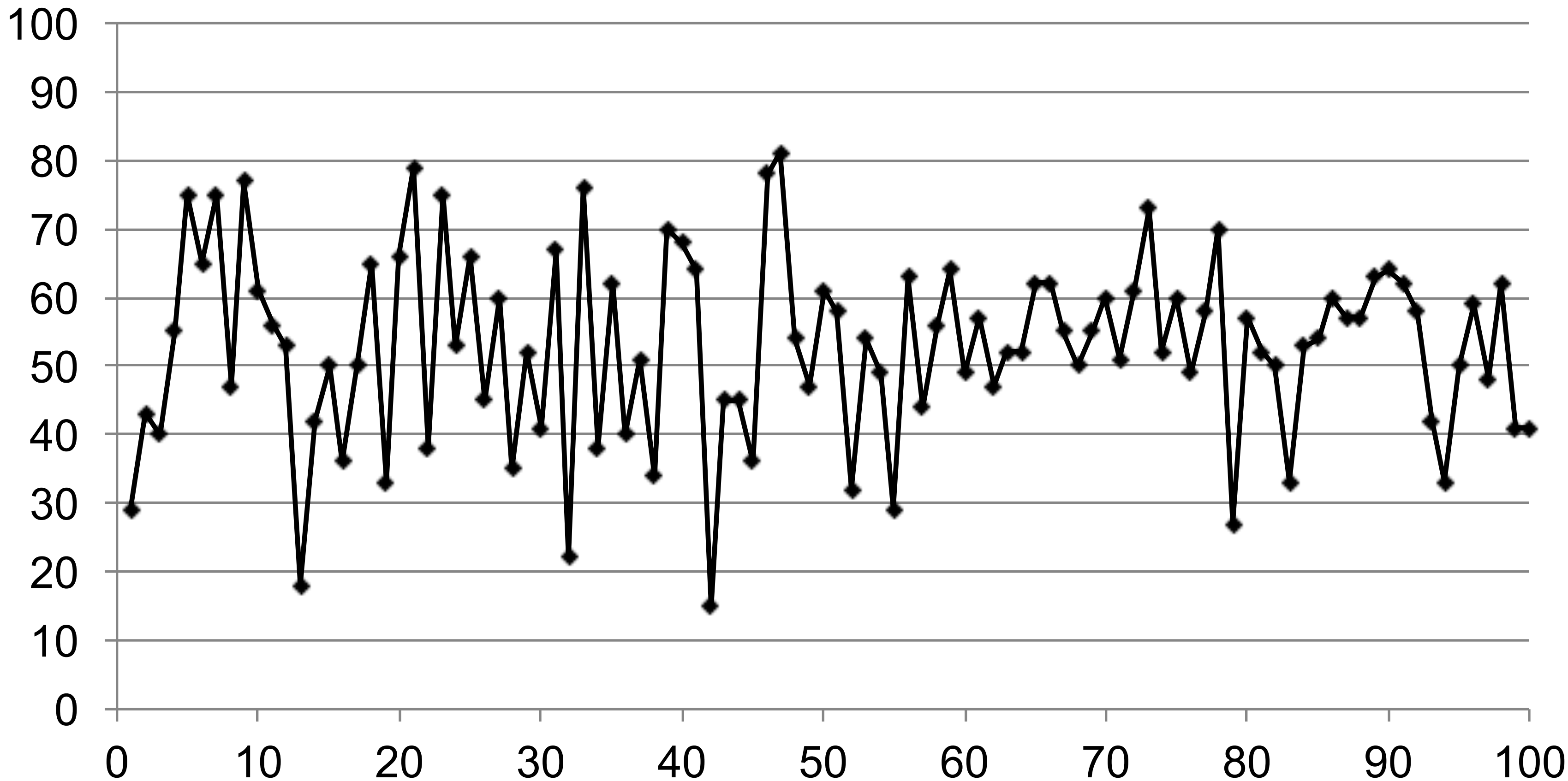
- Forecasting with Past Historical Data
- Moving Averages
- Exponential Smoothing

- Thinking about Trends and Seasonality

Session 3

- Forecasting for New Products
- Fitting distributions

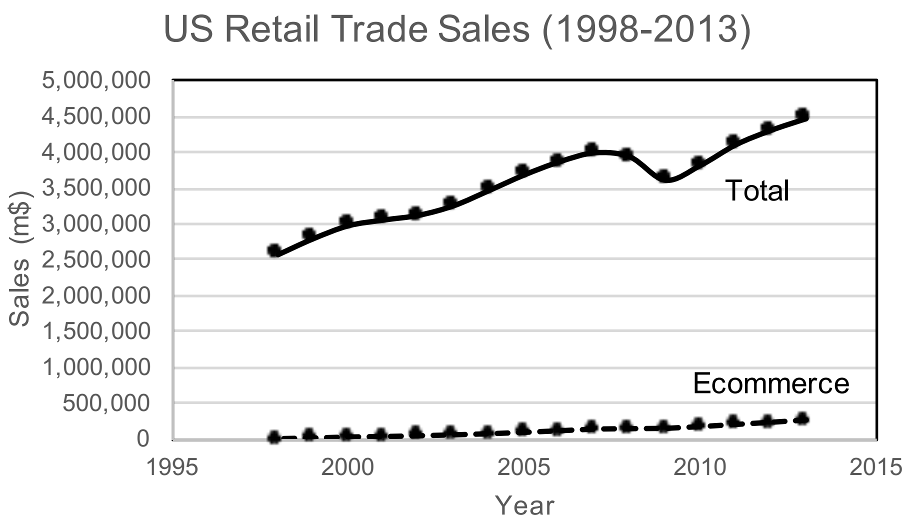
Recall our Newsvendor demand data



- This data is stationary, i.e. demand is largely steady with some noise/variations.
- There is no perceptible trend.

There is often Trend in Data

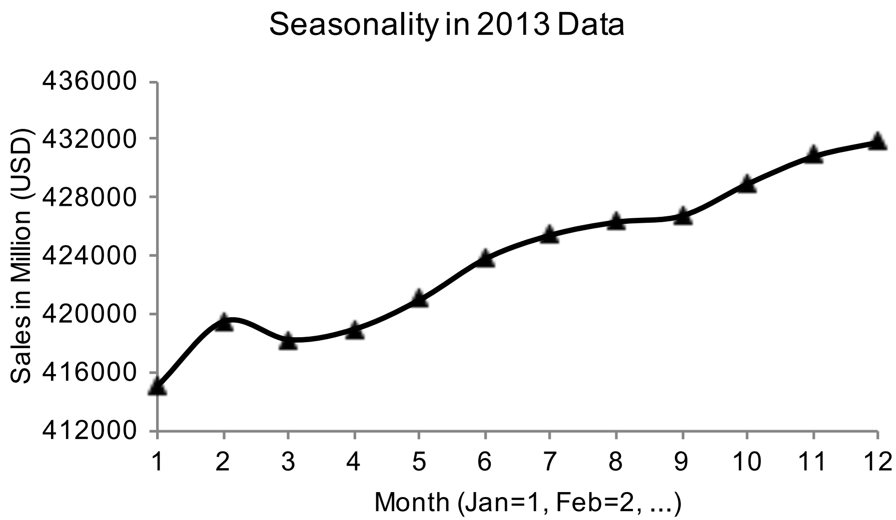
- For example, US Retail Trade Sales - both ecommerce sales and total sales - show a generally increasing trend.



  – Source: census.gov

Data may have seasonality

- For example, some months may consistently have more retail sales than others



- E.g. in 2013, sales are higher during the later half of the year (source: census.gov)
- Seasonality effects often come from predictable annual events (cultural, weather)
  - Thanksgiving sales in the US, Boxing Day sales in Canada, UK and Australia
  - Diwali Sales in India, Chinese New Year Sales.
  - Ski sales in winter.

Forecasting when there is Trend

- In such cases, where the data reveals trend let us examine two ways of forecasting the future demand.

- Using moving averages

- Linear Regression

- In addition, we can also adapt exponential smoothing (advanced material) to adjust for trend.

Moving Averages Lag Trend

- If there is increasing or decreasing trend in data, forecasts generated by moving averages lag behind trend.

- When there is an increasing trend,
  - MA forecasts are usually below the demand.

- When there is a decreasing trend,
  - MA forecasts stay above the demand.

Moving Averages Lag Trend: An illustration

Period	Demand	MA(2)	MA(3)	MA(4)
1	10			
2	20			
3	30	15		
4	40	25	20	
5	50	35	30	25
6	60	45	40	35
7	70	55	50	45
8	80	65	60	55

- Moving Average MA(2) for period 5, is (30+40)/2 = 35.
- Using more data, MA lags behind trend even further
  - MA(3) forecast for period 5, is (20+30+40)/3 = 30.
  - MA(4) forecast for period 5, is (10+20+30+40)/4 = 25.

- How to fix this issue?

Using Regression for Time Series Forecasting

- Main forecasting idea is to fit a line that has a slope to capture the trend in data.

- Linear Regression Methods Can be Used When Trend is Present.

  Model: $D_t = a + bt$

- In this case, the forecast for a period t, is calculated by

  noting the time period t,
  multiplying t by slope b and
  adding the intercept a.

"Best Fit" Trend Line

- Linear Regression

- Ordinary Least Squares (OLS) is method that fits, a trend line through the data *minimizing the squared errors*.

- Mathematically, a straight line $D_t = a + bt$ is fit through the $t = 1, \ldots, n$ data points, and the parameters a and b are chosen to minimize the average squared distance of the data from the trend line.

- Instead of presenting the algebra of how to do fit a trend line on paper,
  - I will present how to fit a trend line in excel using an example.

An Example: Visitors to Yellowstone National Park

Source:nps.gov

- Yellowstone National Park is the most visited National Park in the United States.

- In the File, YellowstoneTemplate.xlsx we have data on the Annual visitors to the National Park (source: nps.gov)

Forecasting visitors to the park

- National Park Service cares about the data for several purposes
  - To plan for camping or backpacking permits.
  - To plan for adequate emergency services.
  - To plan food and fuel services (for visitors and cars).
  - To budget revenues and costs.
  - To understand ecological footprint.

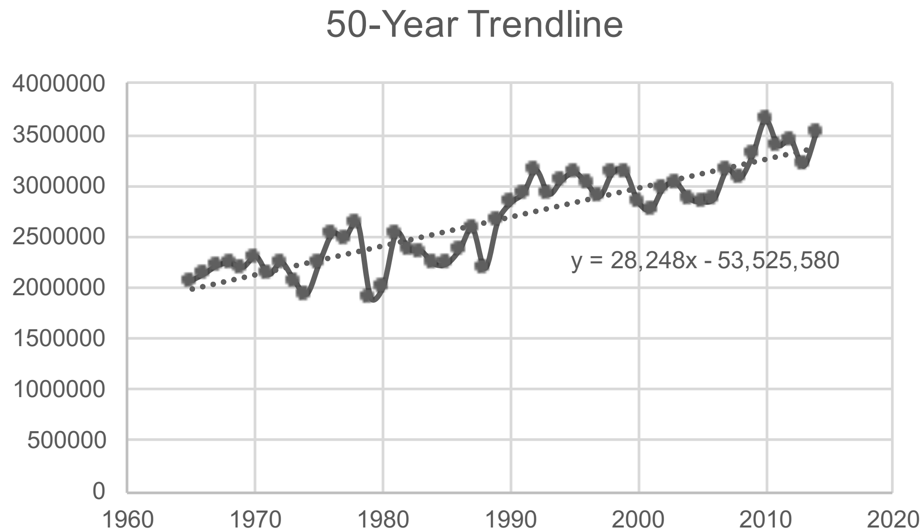
- The data from 1904-2014 shows an increasing trend. We will see how we fit a straight line, to use for forecasting.

- I'll demonstrate how to fit a straight line, using 2 examples
  - the last 50 data points (1965-2014)
  - the last 30 data points (1985-2014).

Fitting Trend line: Example 1

- In YellowstoneTemplate.xlsx template,
- Using last 50 data points we get the following best fit trend line,

$$D_t = -53{,}525{,}580 + 28248t$$

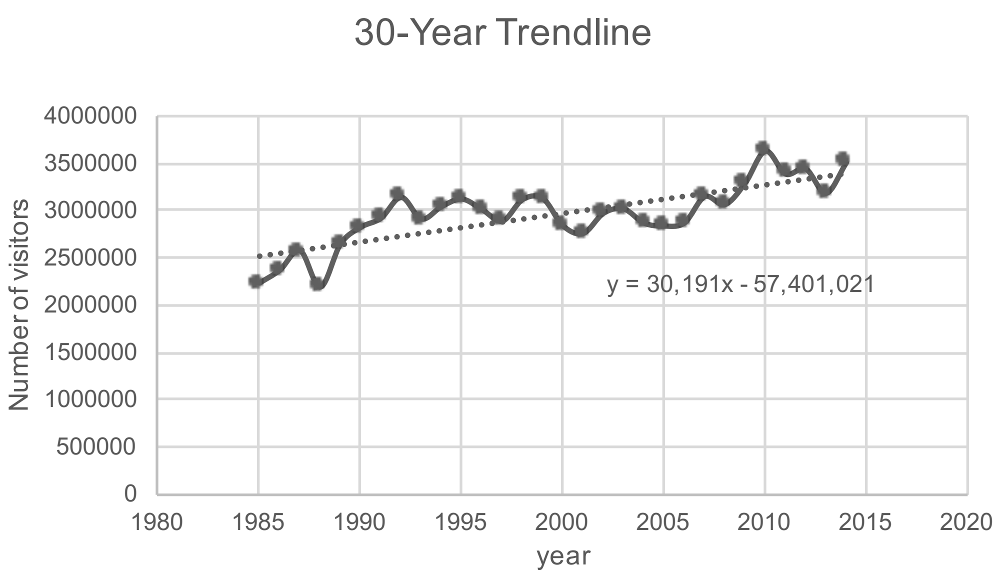


- A forecast for year 2017 would be 3,450,636 visitors.

Trend Line: Example 2

- Using last 30 data points, we get

$$D_t = -57{,}401{,}021 + 30191t$$



- 2017 forecast would be 3,494,226 visitors.
- The generated charts are available in YellowstoneSolution.xlsx

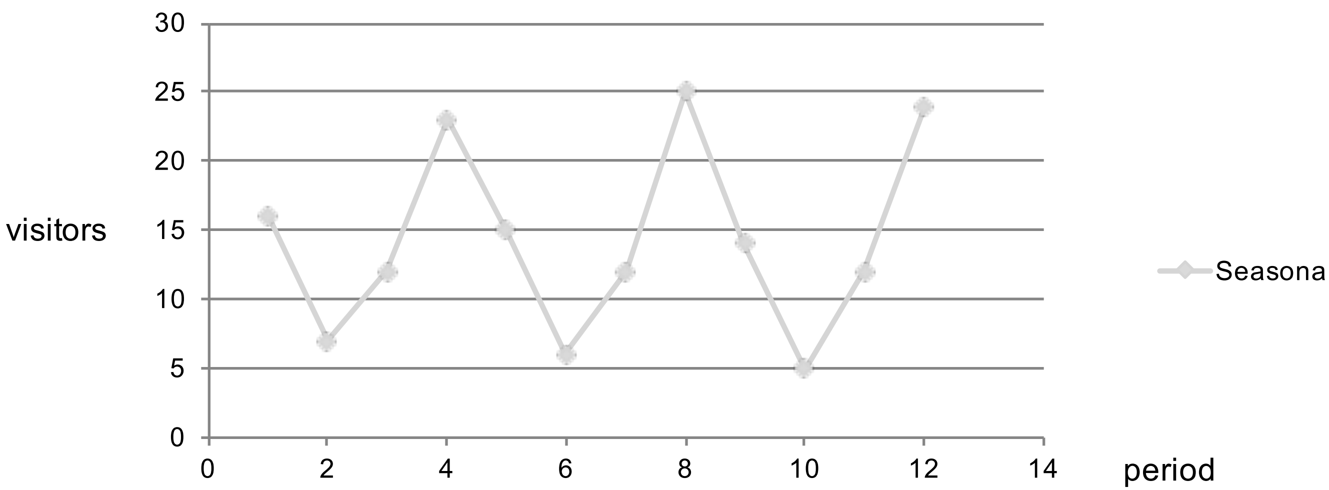
Addressing Seasonality in Data

- Recognizing Seasonality
- Calculating Seasonal factors.

Seasonal Forecasts: An example

Visitors to National Park ('000)

	2012	2013	2014
Fall	16	15	14
Winter	7	6	6
Spring	12	12	12
Summer	23	25	24



Forecasting For Seasonal Series

- Seasonality corresponds to a pattern in the data that repeats at regular intervals.

- Multiplicative seasonal factors c_i (c_1 , c_2 , . . . , c_N) where i = 1 is first season, i = 2 is second season, etc.. N is the total number of seasons.

  – $\sum_i c_i = N$.

  – c_i = 1.25 implies 25% higher than the baseline on average.

  – c_i = 0.75 implies 25% lower than the baseline on average.

- In retail industry, December sales are significant.
  – This means December sales will have a high seasonality factor in sales data.

A Method of Estimating Seasonal Factors

- Step 1: <u>Sample Mean</u>. Compute the sample mean of the entire data set.

- Step 2: <u>Seasonal Averages</u>. Average the observations for the N *like* periods in the data.
  - For example, average all summers, winters, etc.

- Step 3: <u>Seasonal Factors</u>. Divide the averages from Step 2 by the sample mean.
  - The resulting N numbers will exactly add to N and correspond to the N seasonal factors.

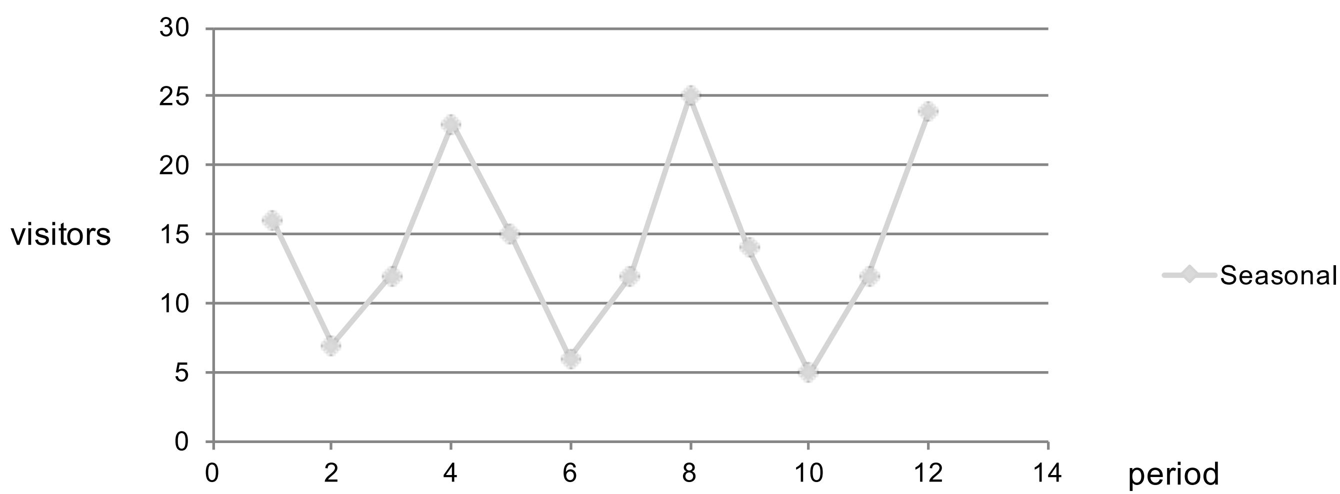
- Step 4: <u>De-seasonalization</u>: To remove seasonality from a series, simply divide each observation in the data by the appropriate seasonal factor.
  - The resulting series will have no seasonality and is called a *de-seasonalized* series.

Example: Step 1 - Calculate Sample Mean

Visitors to National Park ('000)

	2012	2013	2014
Fall	16	15	14
Winter	7	6	6
Spring	12	12	12
Summer	23	25	24

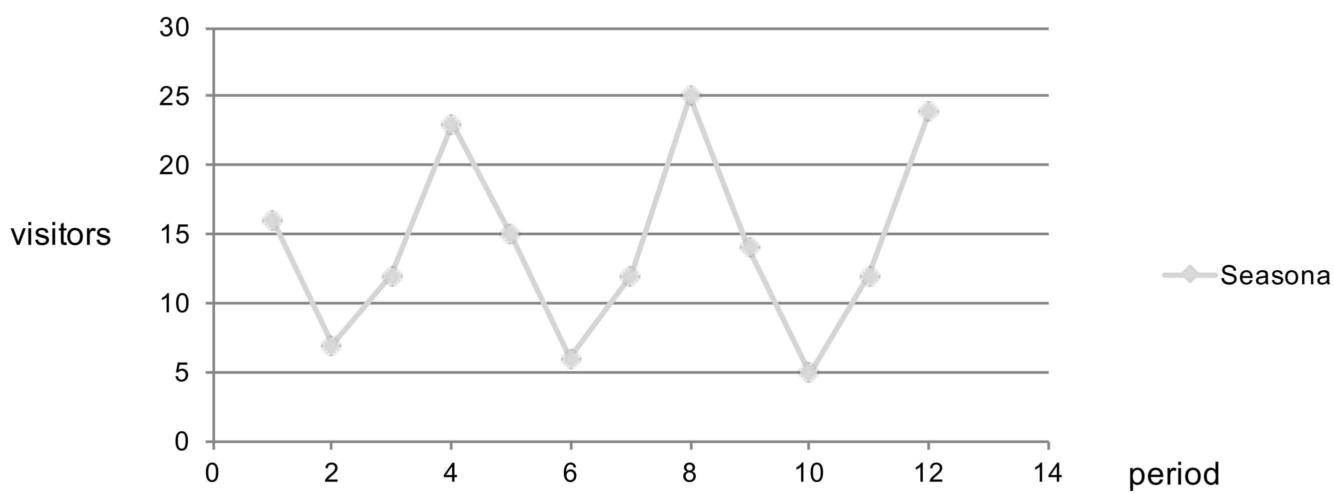
MEAN 14.33



Example: Step 2 – Calculate Season Averages

	Visitors to National Park ('000)			
	2012	2013	2014	
Fall	16	15	14	15
Winter	7	6	6	6.33
Spring	12	12	12	12
Summer	23	25	24	24

MEAN 14.33

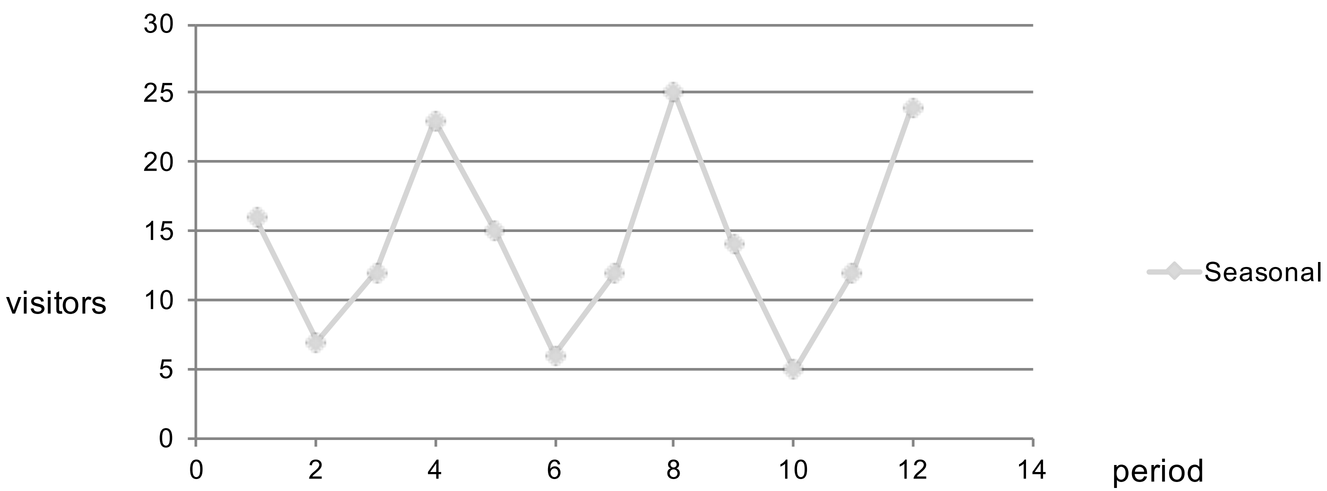


Example: Step 3: Calculate Seasonal Factors

Visitors to National Park ('000)

	2012	2013	2014			c
Fall	16	15	14	15	15/14.33=	**1.05**
Winter	7	6	6	6.33	6.33/14.33=	**0.44**
Spring	12	12	12	12	12/14.33=	**0.84**
Summer	23	25	24	24	24/14.33=	**1.67**

MEAN 14.33



Step 4: De-seasonalized Series

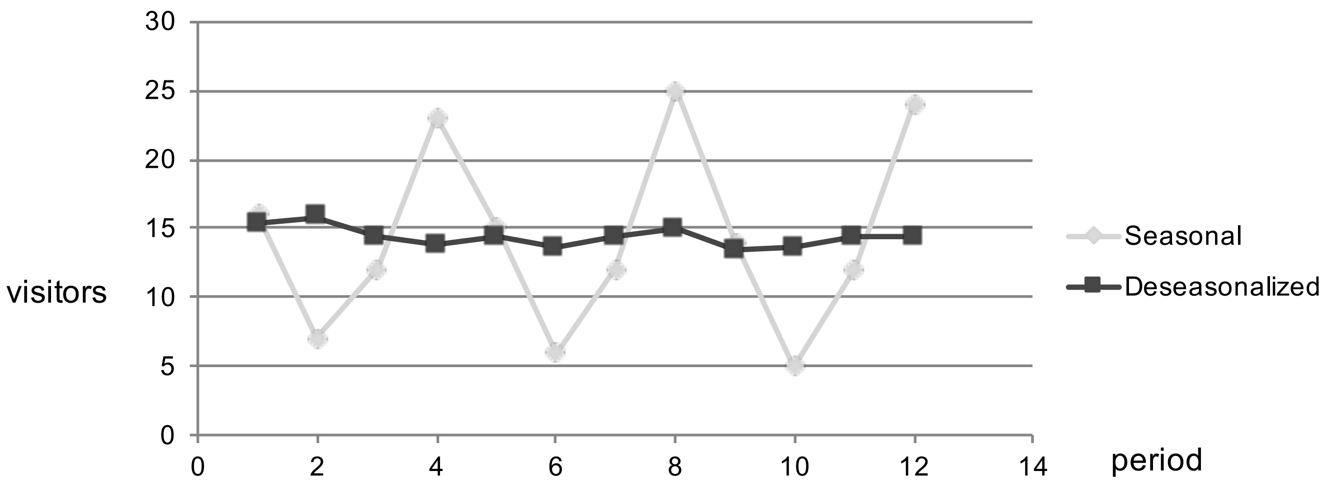
	Visitors to National Park ('000)				c
	2012	2013	2014		
Fall	16	15	14	15	**1.05**
Winter	7	6	6	6.33	**0.44**
Spring	12	12	12	12	**0.84**
Summer	23	25	24	24	**1.67**
		MEAN	14.333		

- Generate de-seasonalized series by dividing each data by the corresponding seasonal factors.

	Visitors to National Park ('000) - deseasonalized			c
	2012	2013	2014	
Fall	16/1.05	15/1.05	14/1.05	**1.05**
Winter	7/0.44	6/0.44	6/0.44	**0.44**
Spring	12/0.84	12/0.84	12/0.84	**0.84**
Summer	23/1.67	25/1.67	24/1.67	**1.67**

Deseasonalized Series

Visitors to National Park ('000) - deseasonalized				c
	2012	2013	2014	
Fall	15.24	14.29	13.33	**1.05**
Winter	15.91	13.64	13.64	**0.44**
Spring	14.29	14.29	14.29	**0.84**
Summer	13.77	14.97	14.37	**1.67**



- The resulting series has no seasonality and is called a *de-seasonalized* series.
  - Can be treated as a stationary series.

Airline Example

- On the website, in Week1AirlineTemplate.xlsx (source: Bureau of Transportation Services).
  - load factors for top 100 US markets is presented.
  - Load factor = (No of passengers / Available seats) *100
  - A load factor of 80% means 80% of the seats are filled.
- The monthly load factor data (2003-2013) shows both trend and seasonality.
  - You can see airlines are getting more crowded over the years.
  - You also see that there some months are more crowded than others.
- Using the Template, I follow the previous 4 steps again to generate the de-seasonalized data.
- All 4 steps are presented in Week1AirlineSolution.xlsx
- I hope you find the excel sheets useful.

Forecasting Seasonal Data

- Estimate seasonal factors and build the de-seasonalized series.
- Forecast is made using the de-seasonalized series, treating it as a stationary series.
  - See lecture material in Session 2.
  - For instance, one could simply use the moving averages method.
- Multiply that forecast by the appropriate seasonal factor to obtain a final forecast.
- I will continue with the example.

Forecasting with Seasonality: Example (contd..)

Visitors to National Park ('000) - deseasonalized				c
	2012	2013	2014	
Fall	15.24	14.29	13.33	**1.05**
Winter	15.91	13.64	13.64	**0.44**
Spring	14.29	14.29	14.29	**0.84**
Summer	13.77	14.97	14.37	**1.67**

- Let's forecast for 2015 Winter using moving average of 4 periods.
- Using MA(4). The average of last 4 data points from year 2014: 13.91
  - De-seasonal forecast = 13.91
- Forecast for winter 2015: De-seasonal Forecast * Winter Seasonal Factor
  - Final Forecast = 13.91 *0.44 = 6.11

A Final Snapshot

	Visitors to National Park ('000)			
	2012	2013	2014	
Fall	16	15	14	15
Winter	7	6	6	6.33
Spring	12	12	12	12
Summer	23	25	24	24
		MEAN	14.333	

- Forecast for winter 2015

  = De-seasonal Forecast * Winter Seasonal Factor

  = 6.11

- Also see the Airline Example: Week1AirlineTemplate.xlsx

Some Thoughts

- Initialization issues exist for any chosen forecasting method.
  - We will examine an idea to address this limited data issue in the next session.
- There is often a model-selection problem in how much and what data to use
  - We will see more of such issues in the upcoming lectures.
- Simple time-series short-term forecasting methods perform well.
  - long term forecasting (assuming same trend, etc.) is fraught with pitfalls since technology changes might occur.
- Tracking of errors is useful for locating forecast bias.
  - We will look at how tracking forecast errors help you model demand.

Next

- An Operational Decision Problem

- Forecasting with Past Historical Data
- Moving Averages
- Exponential Smoothing

- Thinking about Trends and Seasonality

- Forecasting for New Products
- Fitting distributions

Session 4