All handouts for this class: **https://tinyurl.com/IST772crowston**

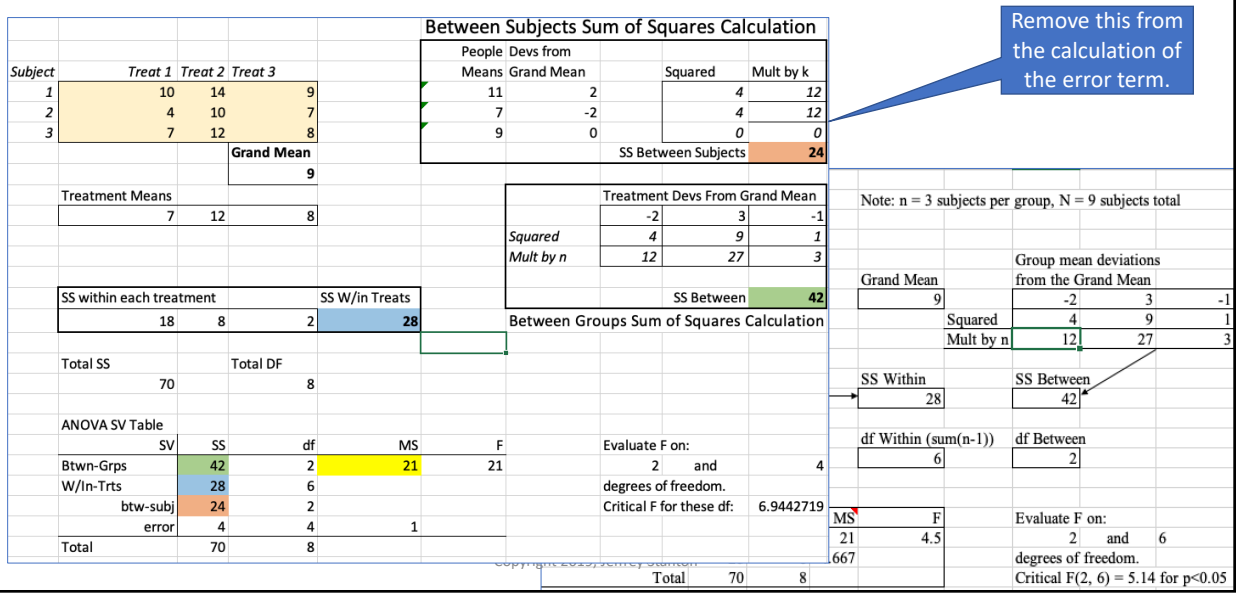
IST772 Analyzing Change (Week 10)

Pre-class activity:

- Grab a copy of 1. ReptMeasuresANOVA.xlsx from the handouts area
- Compare the sources of variance between he two different sheets
- What's the big difference?

1

Within Ss ANOVA vs Between Ss ANOVA



Subject	*Treat 1*	*Treat 2*	*Treat 3*
1	10	14	9
2	4	10	7
3	7	12	8

Grand Mean 9

Treatment Means: 7, 12, 8

SS within each treatment: 18, 8, 2 — SS W/in Treats: **28**

Total SS: 70 — Total DF: 8

ANOVA SV Table

SV	SS	df	MS	F
Btwn-Grps	42	2	21	21
W/In-Trts	28	6		
btw-subj	24	2		
error	4	4	1	
Total	70	8		

Between Subjects Sum of Squares Calculation

People Means	Devs from Grand Mean	Squared	Mult by k
11	2	*4*	*12*
7	-2	*4*	*12*
9	0	*0*	*0*
		SS Between Subjects	**24**

Treatment Devs From Grand Mean			
	-2	3	-1
Squared	*4*	*9*	*1*
Mult by n	*12*	*27*	*3*

SS Between: **42**

Between Groups Sum of Squares Calculation

Evaluate F on: 2 and 4 degrees of freedom. Critical F for these df: 6.9442719

Remove this from the calculation of the error term.

Note: n = 3 subjects per group, N = 9 subjects total

Grand Mean		Group mean deviations from the Grand Mean		
9		-2	3	-1
	Squared	4	9	1
	Mult by n	12	27	3

SS Within	SS Between
28	42

df Within (sum(n-1))	df Between
6	2

MS	F
21	4.5
.667	

Evaluate F on: 2 and 6 degrees of freedom.
Critical F(2, 6) = 5.14 for p<0.05

2

Balancing data

- The approach to balancing in the text includes the command as.numeric, which only works if the IDs are numeric
- Another approach is shown in 2. Balancing data.Rmd
- 2 formats for repeated measure data
  - Long: one row per observation, so multiple rows per subject
  - Wide: one row per subject, with observations in columns
  - We want long for the ANOVA, but it's easier to check completeness in wide

3

Mixed effects models

- A more modern approach to handling repeated measures is "mixed effect models"
  - Separate predictors into "fixed effects" and "random effects"
  - Fixed effects are predictors with values of interest
  - Random effects are predictors with random values that are not themselves of interest (e.g., which chick)
- Advantages
  - Can handle a broader range of models than repeated measures ANOVA (e.g., logistic regression)
  - Can handle missing data
- See mixed-effect-models.pdf in the Google Drive

4

Time series analysis

5

Data Frame vs Time Series

Classes ‘spec_tbl_df’, ‘tbl_df’, ‘tbl’ and 'data.frame': 359 obs. of 6 variables:

$ X1 : num 1 2 3 4 5 6 7 8 9 10 ...

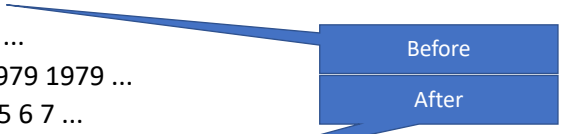
$ year : num 1978 1978 1978 1979 1979 ...

$ month: num 10 11 12 1 2 3 4 5 6 7 ...

$ site1: num 334 337 338 340 341 ...

$ site2: num 332 334 335 336 337 ...

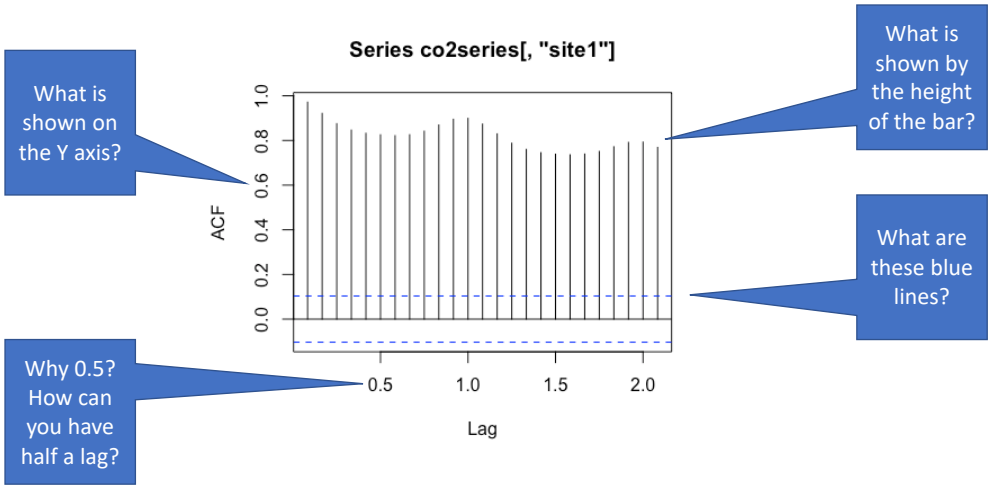
$ site3: num 332 339 339 340 341 ...




```
> co2series <- ts(co2data[,4:6],start=c(1978,10),frequency=12)
> str(co2series) # Structure confirms time series
 Time-Series [1:359, 1:3] from 1979 to 2009: 334 337 338 340 341 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:3] "site1" "site2" "site3"
```


6

Interpreting an ACF



7

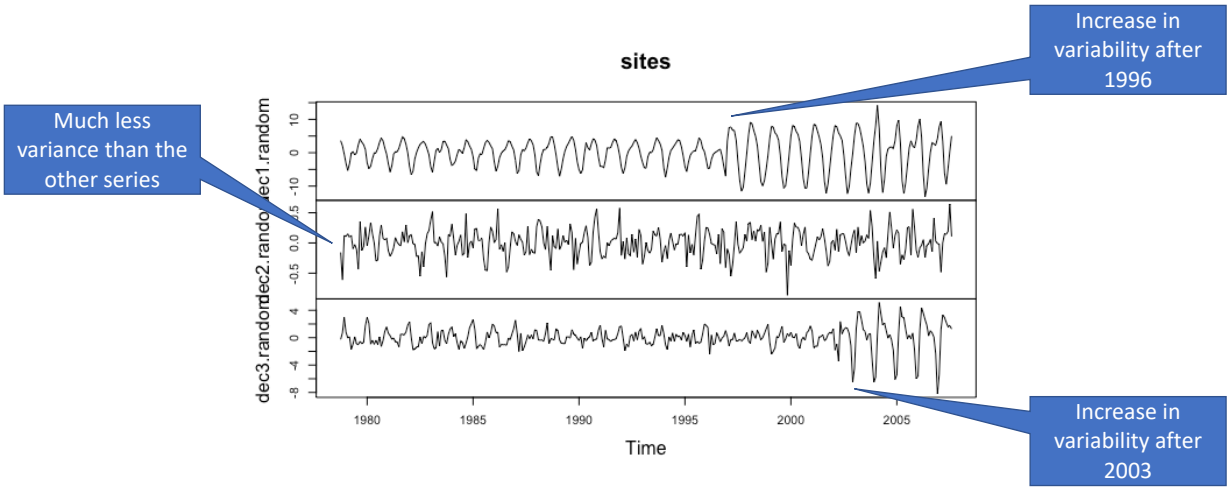
Breakout 1 – Analyze NOAA Data

- Open notebook 3. week_10_time_series_clinic.Rmd
- Read in the time series data
- Run graphics and diagnostics, convert to a time series object
- Decompose the time series and examine stationarity of noise component
- Correlate the stationary time series with each other
- Comment on the result
- Share your code on https://codeshare.io/aJDyRX

All handouts for this class: **https://tinyurl.com/IST772crowston**

8

Decomposition of TS: Error Component



9

adf.test() on Cold Bay, AK

> adf.test(sites[,1]) # If significant, then stationary

Augmented Dickey-Fuller Test

data: sites[, 1]
Dickey-Fuller = -18.823, Lag order = 7, **p-value = 0.01**
alternative hypothesis: stationary

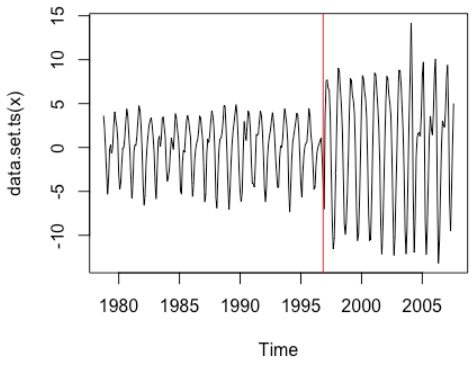
Warning message:
In adf.test(sites[, 1]) : p-value smaller than printed p-value

10

Changepoint Analysis

```
> round(cor(sites),2) # Which two correlate?
             dec1.random dec2.random dec3.random
dec1.random         1.00        0.01       -0.20
dec2.random         0.01        1.00        0.03
dec3.random        -0.20        0.03        1.00
```

Created Using changepoint version 2.2.2
Changepoint type : Change in variance
Method of analysis : AMOC
Test Statistic : Normal
Type of penalty : MBIC with value, 17.54797
Minimum Segment Length : 2
Maximum no. of cpts : 1
Changepoint Locations : 218



11

ARMA in a Nutshell

(1) Autoregressive model of order p (AR(p))

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

i.e., y_t depends on its p previous values

(2) Moving Average model of order q (MA(q))

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q},$$

i.e., y_t depends on q previous random error terms

Image Credit: Moses Johns

12

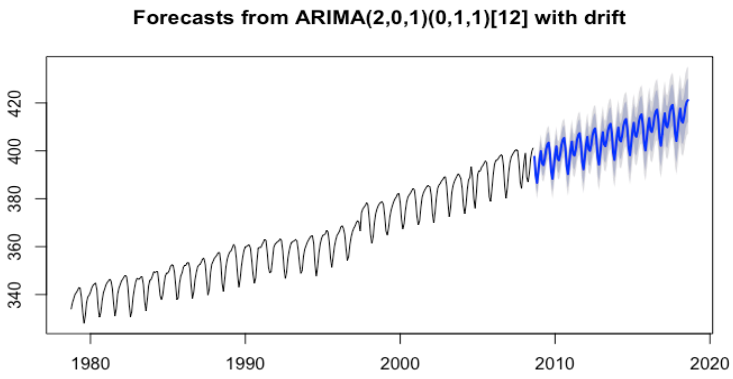
Breakout 3 – Forecasting

- Generate ARIMA models using auto.arima()
- Contained in the forecast package
- Forecasts 10 years of predictions
- Try the "prophet" additive forecasting model created by Facebook researchers

All handouts for this class: **https://tinyurl.com/IST772crowston**

13

Brings us to the Present!



14

Paper of the Week – Hyndman & Khandakar 2007

- Start on Page 14 for discussion of ARIMA
- The essential strategy is to try sequentially several models, trying to minimize AIC or BIC (BIC is Bayesian Information Criterion; both are measures of model error)
- There are hundreds of possible combinations for seasonal models, so the procedure uses short cuts to focus on the models that are most likely to provide good fit
- To reduce the number of models that need to be examined, the procedure tests for "unit roots" (similar to the augmented Dickey-Fuller test for stationarity)

15

Homework and practice exam

- Homework for week 10
  - Is based on exercises 2, 5, 6, 7, and 8 on pages 272 and 273 but with edits in the notebook file
  - Due Saturday 8:30 pm ET
- Third practice exam for the final (final one for the course)
  - Posted in the handouts sharing area soon
  - Submit it to the LMS if you have questions or concerns you'd like addressed
  - I will post a key for the practice exam on Friday

16

Review and final

- Review session
  - Live session plan is to give Class 11 to work on the exam
  - We can do a review session
- Final exam
  - I will send each of you an email with the final exam data and specification after Class 11
  - Exam is due no later than Tuesday 31 March 8:30pm ET

17