Statistically Speaking: Two-way *ANOVA and Repeated Measures*

DAVID L. TABB, PH.D.

OCTOBER 12, 2017

Overview

- Returning to our friend ANOVA, this time with multiple factors in play
- Examining issues that appear in Repeated Measures data sets
- Diagramming some functions used in today's script

Two-way ANOVA

- What if multiple independent variables contribute to the dependent variable?
  - Simple example: subject IQ scores and time spent studying impact final grade
- What if combinations of two factors are required for change in dependent variable?
  - Silly example: desire to host parties and availability of hosting space

Key term: interaction

- If the impact of a level of one factor depends upon the level of another factor.
- With interaction, the “main effects” (impact of individual factors) are not the whole story.
- We must consider three relationships:
  - Impact of A upon dependent variable
  - Impact of B upon dependent variable
  - Impact of A*B upon dependent variable

Assumptions for all ANOVAs

- The observations are independent of each other: are any subjects related?
- The variation in each set is normally distributed. ANOVA tolerates outliers badly.
- Sets have homogenous variance. (We call this type of set *homoscedastic*.)

Repeated measures

- Many studies collect multiple measurements for an individual.
- Time series: an individual produces a datum at each of several time points
  - "We measured at 1, 2, and 6 months."
- Spatial set: an individual produces data for several related sites
  - "Each patient gave serum and saliva."

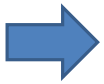
Subjects and variation

- Between-subject variation: If subjects take on only one level of a factor, it is described as "between subjects." For example, a child falls in either the 8yo or 12yo group, not both.
- Within-subject variation: If each subject appears at each level of a factor, it is described as "within a subject." For example, each child is tested in all five different tasks.

Lost in time, lost in space

- Clustered: Different samples group together.
- Repeated Measures (ANOVA):
  - Measurements may be separated by location rather than time.
  - More measurements implies shift from categorical to continuous handling.
- Longitudinal Data (mixed effects):
  - "Dropouts" frequently complicate analysis.
  - Autocorrelation useful for variable intervals.

Instruction	Student	Month	Calories.per.day
Curriculum A	a	1	2000
Curriculum A	a	2	1978
Curriculum A	a	3	1962
Curriculum A	a	4	1873
Curriculum A	a	5	1782
Curriculum A	a	6	1737
Curriculum A	b	1	1900
Curriculum A	b	2	1826
Curriculum A	b	3	1782
Curriculum A	b	4	1718
Curriculum A	b	5	1639
Curriculum A	b	6	1644
Curriculum A	c	1	2100
Curriculum A	c	2	2067
Curriculum A	c	3	2065
Curriculum A	c	4	2015
Curriculum A	c	5	1994
Curriculum A	c	6	1919
Curriculum A	d	1	2000
Curriculum A	d	2	1981
Curriculum A	d	3	1987
Curriculum A	d	4	2016
Curriculum A	d	5	2010
Curriculum A	d	6	1946
Curriculum B	e	1	2100
Curriculum B	e	2	2004
Curriculum B	e	3	2027
Curriculum B	e	4	2109
Curriculum B	e	5	2197
Curriculum B	e	6	2294
Curriculum B	f	1	2000
Curriculum B	f	2	2011
Curriculum B	f	3	2089
Curriculum B	f	4	2124
Curriculum B	f	5	2199
Curriculum B	f	6	2234
Curriculum B	g	1	2000
Curriculum B	g	2	2074
Curriculum B	g	3	2141
Curriculum B	g	4	2199
Curriculum B	g	5	2265
Curriculum B	g	6	2254
Curriculum B	h	1	2000
Curriculum B	h	2	1970
Curriculum B	h	3	1951
Curriculum B	h	4	1981
Curriculum B	h	5	1987
Curriculum B	h	6	1969
Curriculum C	i	1	1950
Curriculum C	i	2	2007
Curriculum C	i	3	1978
Curriculum C	i	4	1965
Curriculum C	i	5	1984
Curriculum C	i	6	2020
Curriculum C	j	1	2000
Curriculum C	j	2	2029
Curriculum C	j	3	2033
Curriculum C	j	4	2050
Curriculum C	j	5	2001
Curriculum C	j	6	1988
Curriculum C	k	1	2000
Curriculum C	k	2	1976
Curriculum C	k	3	2025
Curriculum C	k	4	2047
Curriculum C	k	5	2033
Curriculum C	k	6	1984
Curriculum C	l	1	2000
Curriculum C	l	2	2020
Curriculum C	l	3	2009
Curriculum C	l	4	2017
Curriculum C	l	5	1989
Curriculum C	l	6	2020



Instruction	Student	Month	Calories.per.day
Curriculum A	a	1	2000
Curriculum A	a	2	1978
Curriculum A	a	3	1962
Curriculum A	a	4	1873

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Deep or Wide? Reshape!

Instruction	Student	Month1	Month2	Month3	Month4	Month5	Month6
Curriculum A	a	2000	1978	1962	1873	1782	1737
Curriculum A	b	1900	1826	1782	1718	1639	1644
Curriculum A	c	2100	2067	2065	2015	1994	1919
Curriculum A	d	2000	1981	1987	2016	2010	1946
Curriculum B	e	2100	2004	2027	2109	2197	2294
Curriculum B	f	2000	2011	2089	2124	2199	2234
Curriculum B	g	2000	2074	2141	2199	2265	2254
Curriculum B	h	2000	1970	1951	1981	1987	1969
Curriculum C	i	1950	2007	1978	1965	1984	2020
Curriculum C	j	2000	2029	2033	2050	2001	1988
Curriculum C	k	2000	1976	2025	2047	2033	1984
Curriculum C	l	2000	2020	2009	2017	1989	2020

http://seananderson.ca/2013/10/19/reshape.html

`switch` statement

- Serum is a factor taking on three levels: ‘A’, ‘B’, or ‘C’. They are introduced in this order.
- A particular item in Serum can be used to decide the return value for switch function:

```
Return=switch(Serum[8],3.0,1.0,2.0)
```

for loop

- We need to *iterate* through a group of instructions for each member in a set.
- We want to print the square of each number from 1:10.

```
for (looper in 1:10) {
  cat(looper^2,"\n")
}
```

Finish this line, and go to the next.

Indexing

- We have a vector of numbers. We want to look at the *n*th value. `Vector[n]`
- From a matrix of numbers, we want the item from the *i*th row and the *j*th column. `Matrix[i,j]`
- We have a data frame of several vectors. We want to grab the Date field. `DF$Date`

See also: subset() function.

Takeaways

- Two-way ANOVA can reveal both main effects and interactions of factors.
- ANOVA of repeated measures sets work if one can produce a suitable model. Note that long vectors for each subject are better handled through mixed effects models.
- R features functions for loops and logic, just like other languages.