

ANOVA insights

Let's revisit an example from a previous class where we looked at data showing the number of job offers that a sample of students had at the time of graduation, together with their major field of study

Observations		
DS	Bio	Econ
2	3	2
5	0	2
1	1	0
	2	
$n = 3$	$n = 4$	$n = 3$
$\bar{y} = 2.67$	$\bar{y} = 1.5$	$\bar{y} = 1.33$

One way to think about each data value is as follows

$$\text{Observation} = \text{Grand mean} + \text{Factor effect} + \text{Residual}$$

Mathematically, this can be written as

$$y_i = (\bar{y}) + (\bar{y} - \bar{y}) + (y_i - \bar{y})$$

where \bar{y} = Grand mean, \bar{y} = group mean, y_i = data value (observation)

Exercise:

1. Write the first couple of observations in the job offers data in this form.
2. Show that the general form of the equation given above always holds.
3. Explain each term in the parentheses in a practically useful way.
E.g., each data value is, essentially, the sum of ...

ANOVA insights (continued)

Here's the equation again

$$\text{Observation} = \text{Grand mean} + \text{Factor effect} + \text{Residual}$$

Let's make a separate table for each term in the equation:

Observations		
DS	Bio	Econ
2	3	2
5	0	2
1	1	0
	2	
$\bar{y} = 2.67$	$\bar{y} = 1.5$	$\bar{y} = 1.33$

Grand mean			Factor effect			Residual		
DS	Bio	Econ	DS	Bio	Econ	DS	Bio	Econ
1.8	1.8	1.8	0.87	-0.3	-0.47	-0.67	1.5	0.67
1.8	1.8	1.8	0.87	-0.3	-0.47	2.33	-1.5	0.67
1.8	1.8	1.8	0.87	-0.3	-0.47	-1.67	-0.5	-1.33
	1.8			-0.3			0.5	

$$\sum 1.8^2 = 32.4$$

$$\sum \text{all}^2 = 3.26667$$

$$\sum \text{all}^2 = 16.33333$$

Exercise:

1. Can you identify SS_B and SS_W in the above sums?
2. Add the 3 sums given above
3. Compute $\sum (y_i)^2$ for the data values (i.e., in the 1st table).

Moral of the story?!

It turns out that

$$\sum \text{Observation}^2 = \sum \text{Grand mean}^2 + \sum \text{Factor effect}^2 + \sum \text{Residual}^2$$