NOTES ON ANALYSIS OF VARIANCE (ANOVA)

Statement

- The analysis of variance (ANOVA) is a statistical technique that allows to decompose the sum of squares of a response variable into components attributable to known sources of variability (independent factors) and unknown suorces of variability (error).
- ANOVA separates the effects of factors in order to measure and judge their relative values.
- The application of ANOVA regards quantitative variables normally distributed.

Example no.1

ADG (hg/d) in 3 groups ($i=1,\ldots,3$) of piglets treated with 3 different levels ($j=1,\ldots,3$) of B12 vitamin

	B12 (mg)		
Group	0	5	10
1	1.04	1.52	1.63
2	1.00	1.56	1.57
3	0.69	1.54	1.54

Example no. 1: sums & means

	B12 (mg)			General
Group	0	5	10	
1	1.04	1.52	1.63	
2	1.00	1.56	1.57	
3	0.69	1.54	1.54	
Sum	2.73	4.62	4.74	12.09
Mean	0.91	1.54	1.58	1.34

Example no.1: sum of squares

Total sum of squares: TSS (ADG di 9 groups of piglets)

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\overline{y}\right)^{2}=\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}^{2}-\frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}\right)^{2}}{n}=\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}^{2}-\frac{\overline{y}^{2}}{n}$$

Mean=(1.04+1.00+0.69+1.52+1.54+1.63+1.57+1.54)/9=1.34

1) TSS=[$(1.04-1.34)^2+(1.00-1.34)^2+$........$+(1.54-1.34)^2$]=0.93

2) TSS=($1.04^2+1.00^2+$......$+1.54^2$)-($1.34^2/9$)= 0.93

With $1.34^2/9$= Correction Factor (CF)

Decomposition of Total Sum of Squares - 1

Differences $$\left(y_{ij}-\bar{y}\right)=\left(y_{ij}-\bar{y}_{i}\right)+\left(\bar{y}_{i}-\bar{y}\right)$$

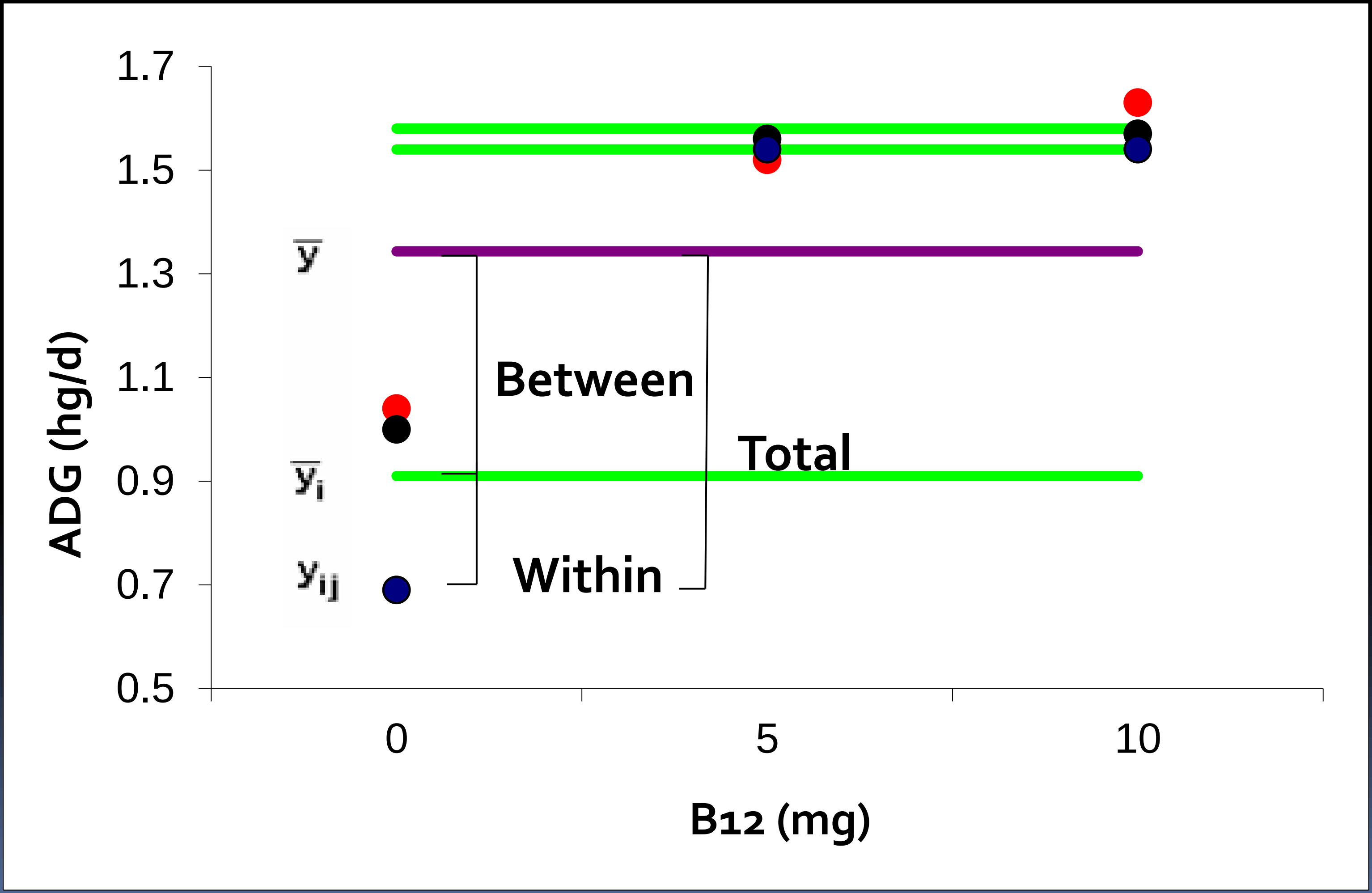
Total Within Between

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\bar{y}\right)^{2}=\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\bar{y}_{i}\right)^{2}+J\sum_{i=1}^{I}\left(\bar{y}_{i}-\bar{y}\right)^{2}$$

Sum of squares: TSS = WSS+BSS

Degrees of freedom (df) n-1 =(n-I) + (I-1)

Decomposition of Total Sum of Squares - 2



Decomposition of Total Sum of Squares - 3

TSS = BSS + WSS

TSS: Total sum of squares

SSB: Between groups sum of squares or sum of squares attributable to the factor(s)

SSW: Error or residual Sum of Squares (within groups SS)

Source of variation	SS	df	MS
Factor	$J\sum_{i=1}^{I}\left(\overline{y}_i-\overline{y}\right)^2$	I-1	BSS/(I-1)
Error	$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\overline{y}_i\right)^2$	n-I	WSS/(n-I)
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\overline{y}\right)^2$	n-1	TSS/(n-1)

Degrees of freedom

- The degrees of freedom are the linearly independent differences of the variable from the mean and usually it os expressed as n-1.
- Since $\Sigma(x_i\text{-}x)=0$,

  Differences from the mean are all independent excep one that assume the value for which the sum to the others produce zero.
- In the ANOVA, the degrees of freedom are calculated considering the independent differences from each mean, i.e. differences from the general mean (n-1), differences between the mean of each level and the general mean (I-1) and differences from the mean of each level (within), that is n-I

Models

- Models are mathematical instruments that allows the representation of cause-effect relationship between variables or groups of variables and factors.
- ANOVA models are stochastics linear models, i.e., characterized by additive effects and accounting for one or more unknown random effect.

$$y_{ij} = \mu + a_i + e_{ij}$$

y_{ij}: single observation on j-th unit of i-th treatment group;
μ: general mean;
a_i: $m-m_i$ effect of i-th treatment level ($i=1, \ldots, I$);
e_{ij}: residual error $N(0,s^2_e)$

Aims of ANOVA

- To test the following hypotheses:

  $$H_o: \alpha_i = o \;\; => \mu=\mu_i$$

  $$H_1: \alpha_i \# o$$

  Verify if means of each treatment are mutually different and different form the general mean with a given probability.
  In other words, the aim of ANOVA is to test if the treatment generate significant effect on the general mean (H_1 is true), or if the means of treatment(s) are not different and equal to the general mean.
- Be careful that the ANOVA test do not discriminate within the treatment groups (between means)
- ANOVA encompasses both estimates of the model parameters and the estimates of variance components

Hypothesis test

- The test function in ANOVA is:

  F = BMS/WMS

  The aim is verify if BMS is bigger than WMS (or residual variance), or to analyze if the between group variance is greater than the within group variance.
  if BMS > WMS => the causes that produce differences between the means of the treatments are more relevant than those due to the error (H_1 is true).
  if BMS <= WMS => the treatment do not produce differences between the means of treatments (H_0 is true).
- F is distributed as Snedecor's F, with (I-1) numerator df and (n-I) denominator df.

Comparison and contrasts

The rejection of the null hypothesis is not always a sufficient result.

Indeed, numerous alternative hypotheses can be taken into account depending on the structure of ANOVA or by the aim of the experiment.

Alternative (even complex) hypotheses are formed by combining means.

Contrasts

A “Contrast” consists in the estimate od a linear function of means belonging to the i-th treatment each multiply by a weight coefficient “c_i”, such that $\Sigma c_i=0$.

$$L(k) = c_1 \bar{y}_1 + c_2 \bar{y}_2 + \ldots\ldots\ldots + c_i \bar{y}_i = \sum_i c_i \bar{y}_i$$

In order to maintain a sufficient and constant probability level for all the contrasts/tests (equal to α), contrasts must be mutually independent, i.e., they are orthogonal.

Orthogonal contrasts

- Contrasts are defined orthogonal if the following rule is respected, i.e. the sum of products between coefficients belonging to row i and column k is zero.

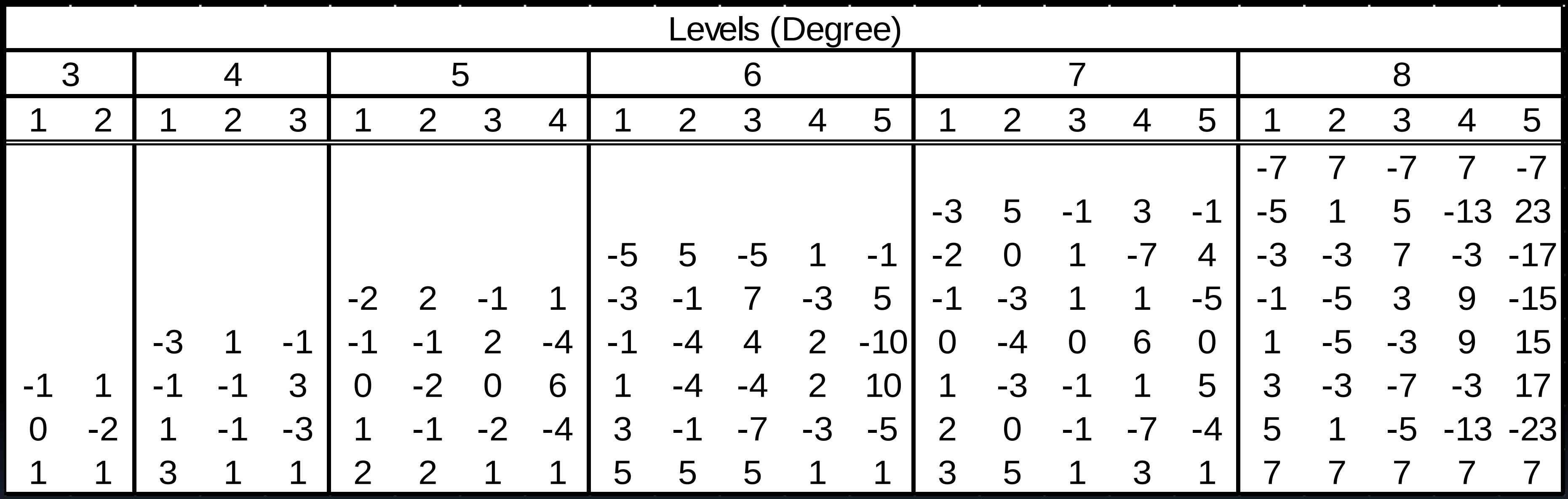
$$\Sigma c_{ik}c_{ik'}=0$$

- Orthogonal contrasts account all “controlled” experimental variability under the null hypotheses that they are not correlated, that is to say they account for independent amount sum of squares due to treatment.

SAS Example

```
data adgpiglets;
input group vitB12 AdG;
cards;
1 0 1.04
2 0 1.00
3 0 0.69
1 5 1.52
2 5 1.56
3 5 1.54
1 10 1.63
2 10 1.57
3 10 1.54
run;
proc glm data= adgpiglets;
class vitB12;
model adg=vitB12;
lsmeans vitB12;
output out=adgpiglets1 p=pred r=resid;
contrast 'contr vs vitb12' vitb12 -2 1 1;
contrast '5 vs 10' vitb12 0 1 -1;
run;
quit;
```

Polynomial coefficients



Levels (Degree)																							
3		4			5				6					7					8				
1	2	1	2	3	1	2	3	4	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
																			-7	7	-7	7	-7
														-3	5	-1	3	-1	-5	1	5	-13	23
									-5	5	-5	1	-1	-2	0	1	-7	4	-3	-3	7	-3	-17
					-2	2	-1	1	-3	-1	7	-3	5	-1	-3	1	1	-5	-1	-5	3	9	-15
		-3	1	-1	-1	-1	2	-4	-1	-4	4	2	-10	0	-4	0	6	0	1	-5	-3	9	15
-1	1	-1	-1	3	0	-2	0	6	1	-4	-4	2	10	1	-3	-1	1	5	3	-3	-7	-3	17
0	-2	1	-1	-3	1	-1	-2	-4	3	-1	-7	-3	-5	2	0	-1	-7	-4	5	1	-5	-13	-23
1	1	3	1	1	2	2	1	1	5	5	5	1	1	3	5	1	3	1	7	7	7	7	7

ANOVA with two or more classification factors

- In this case two or more classification factors (known source of variability) and their interaction are taken into account aiming at:
  - Analyzing the relative weight of two o more source of variation that act in conjunction on experimental units (i.e., variables), as for example 2 drugs at different levels each
  - Separating from residual variance other systematic and known effects that would overestimate the residual variance of not accounted
- It is important to notice that in order to estimate interaction(s), at least two replicates should be considered within each combination of interacting factors

Two way ANOVA model with interaction

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{ijk}$$

$i = 1, \ldots, I$

$j = 1, \ldots, J$

$k = 1, \ldots, K$

$n = I*J*K$

If no interaction A*B could be detected, the 2 analyzed factors are independent.

If one factor produce a different than expected response within the different level of the other factor, it means that interaction of A*B exists, i.e., it is significant

Two way ANOVA with interaction

Source of Variation	Sum of squares	d.f.	Mean Squares	F
Factor A	SSA	I-1	SSA/(I-1)	MSA/MSE
Factor B	SSB	J-1	SSB/(J-1)	MSB/MSE
Interactio AxB	SSAB	(I-1)(J-1)	SSAB/(I-1)(J-1)	MSAB/MSE
Error	SSE	IJ(K-1)	SSE/IJ(K-1)	
Total	SSY	n-1		