Topic 4. Orthogonal contrasts [ST&D p. 183]

ANOVA: H_0: $\mu_1 = \mu_2 = ... = \mu_k$ vs. H_1: at least one different.

If F is significant: Which are the different ones?

The different treatments can be identified by further **partitioning the treatment sum of squares** to provide additional F tests to answer as many **independent planned** questions as there are degrees of freedom for treatments in the ANOVA.

The **contrast** or **orthogonal** approach to mean separation requires an ***a priori*** knowledge, either based on biological considerations or on the results of preliminary experimentation. These contrasts are called **planned F tests** or **single degree of freedom tests**

Sum of SST for each contrast = SST for original experiment.

This means that the experiment can be in effect partitioned into t - 1 separate independent experiments, one for each contrast.

4. 1. Definition of contrast and orthogonality [ST&D p. 183-188]

A **contrast** is a linear sum of the form

$$Q = \sum_{i=1}^{m} c_i \overline{Y}_{i.}; with: \sum_{i=1}^{m} c_i = 0$$

$\overline{Y}_{i.}$ = the treatment means, and m ≤ t, is the number of treatments.

c_i's are usually integers. A contrast always has a single degree of freedom.

Suppose $Q_c = \sum_{i=1}^{m} c_i \overline{Y}_{i.};\ and\ Q_d = \sum_{i=1}^{m} d_i \overline{Y}_{i.}$ are **two contrasts**. Then they are **orthogonal** if the sum of the products of the corresponding coefficients of any two comparisons is zero.

$$\sum_{i=1}^{m} c_i d_i = 0$$

Example

Treatments, T1, T2 and **T3 (control)**. Two d.f. for treatments.

One could test the hypotheses that T1 and T2 are not significantly different from the control: $\mu_1 = \mu_3$ and $\mu_2 = \mu_3$.

$\mu_1 = \mu_3$ $(1\mu_1 + 0\mu_2 - 1\mu_3 = 0)$ the coefficients are: $c_1 = 1$, $c_2 = 0$, $c_3 = -1$
$\mu_2 = \mu_3$ $(0\mu_1 + 1\mu_2 - 1\mu_3 = 0)$ the coefficients are: $d_1 = 0$, $d_2 = 1$, $d_3 = -1$

These linear combinations of means are **contrast** because

$\sum_{i=1}^{m} c_i = 0$ $(1 + 0 + (-1) = 0)$

$\sum_{i=1}^{m} d_i = 0$ $(0 + 1 + (-1) = 0)$

These two contrasts are **not orthogonal** because

$\sum_{i=1}^{m} c_i d_i \neq 0$ $(c_1d_1 + c_2d_2 + c_3d_3 = 0 + 0 + 1 = 1)$.

Not every pair of hypotheses can be tested using this approach.

If we test:

1) The average of T_1 and T_2 is not significantly different from the control,

   $c_1 = 1$, $c_2 = 1$, $c_3 = -2$.

2) μ_1 is not significantly different from μ_2.

   $d_1 = 1$, $d_2 = -1$, $d_3 = 0$

These are **contrasts** since: $1 + 1 + (-2) = 0$ and $1 + (-1) + 0 = 0$

and are **orthogonal** because: $c_1d_1 + c_2d_2 + c_3d_3 = 1 + (-1) + 0 = 0$.

We will discuss two general kinds of linear combinations:

class comparisons and **trend comparisons**.

4. 2. Class comparisons: ANOVA on groups, or *classes.*

Example: Results (mg dry weight) of an experiment (CRD) to determine the effect of seed treatment by acids on the early growth of rice seedlings.

Table 4.1.

Treatment	Replications					Total $Y_{i.}$	Mean $\bar{Y}_{i.}$
Control	4.23	4.38	4.1	3.99	4.25	20.95	4.19
HCl	3.85	3.78	3.91	3.94	3.86	19.34	3.87
Propionic	3.75	3.65	3.82	3.69	3.73	18.64	3.73
Butyric	3.66	3.67	3.62	3.54	3.71	18.2	3.64
Overall						$Y_{..}$ = 77.13	$\bar{Y}_{..}$ = 3.86

Table 4.2. ANOVA of data in Table 4.1.

Source of Variation	df	Sum of Squares	Mean Squares	F
Total	19	1.0113		
Treatment	3	0.8738	0.2912	33.87
Exp. error	16	0.1376	0.0086	

Questions: 1) Do acid treatments decrease seedling growth?
2) Are organic acids different from inorganic acids?
3) Is there a ≠ in the effects of the 2 organic acids?

Table 4.3. Orthogonal coefficients.

		Control	HCl	Propionic	Butyric
	Totals	20.95	19.34	18.64	18.2
Comparisons	Means	4.19	3.87	3.73	3.64
Control vs. acid		+3	-1	-1	-1
Inorg. vs. org.		0	-2	+1	+1
Between org.		0	0	+1	-1

Rules to construct coefficients for class comparisons

1. In comparing the means of **two groups**, each containing the **same number** of treatments, assign **+1** to the members of one group and **-1** to the members of the other. (Example: "Between org.").

2. In comparing groups containing **different numbers** of treatments, assign:

1st group coefficients= number of treatments in the second group

2nd group coefficients= number of treatments in the first group, with opposite sign.

Example: If among 5 treatments, the first two are to be compared to the last three, the coefficients would be +3, +3, -2, -2, -2. (e.g. control vs acids)

3. The coefficients for any comparison should be **reduced to the smallest** possible integers for each calculation. Thus, +4, +4, -2, -2, -2, -2. should be reduced to +2, +2, -1, -1, -1, -1.

4. At times, a comparison component may be an **interaction** of two other comparisons. The coefficients for this comparison are determined by **multiplying the corresponding coefficients** of the two comparisons

Example: Experiment with 4 treatments, 2 levels of N and 2 levels of P.

	N_0P_0	N_0P_1	N_1P_0	N_1P_1
Between N	-1	-1	1	1
Between P	-1	1	-1	1
Interaction (NxP)	1	-1	-1	1

If comparisons are **orthogonal,** the conclusion drawn for one comparison is **independent** of (not influenced by) the others.

COMPUTATION

Sum of squares for a single degree of freedom *F* test for linear combinations of ***treatment means***

$$SS(Q) = MS(Q) = \frac{(\sum c_i \overline{Y}_{i.})^2}{(\sum c_i^2)/r} \; or \; \frac{(\sum c_i \overline{Y}_{i.})^2}{\sum (c_i^2 / r_i)} \; for\ unbalanced\ designs$$

SS_1 (control vs. acid) = $[3(4.19) – 3.64 – 3.73 – 3.87]^2$ / [(12)/5] = 0.74

SS_1 (Inorg. vs. org.) = $[3.64 + 3.73 – 2(3.87)]^2$ / [(6)/5] = 0.11

SS_1 (between org.) = $[-3.64 + 3.73]^2$ / [(2)/5] = 0.02

ST&D≠: Q formulas p. 184: for ***treatment totals*** ($r*\sum c_i^2$), not ***treatment means***.

Table 4.5. Orthogonal partitioning of treatments of Table 4.2.

Source of Var.	df	SS	MS	F
Total	19	1.01		
Treatment	**3**	**0.87**	0.2912	33.87 **
Control vs. acid	**1**	**0.74**	0.7415	86.22 **
Inorg. vs. Org.	**1**	**0.11**	0.1129	13.13 **
Between Org.	**1**	**0.02**	0.0194	2.26 NS
Error	16	0.14	0.0086	

We conclude that

- All three acids significantly reduce seedling growth ($P<0.01$)
- Organic acids cause more reduction than the inorganic acid ($P<0.01$)
- The difference between the organic acids is not significant ($P>0.05$).

When the individual comparisons are orthogonal:

- **SS of contrasts add up to the SST**
- **The maximum number of orthogonal comparisons is t-1**
- **The SS for one comparison does not contain any part of the SS of another comparison.**
- **The conclusions are independent from each other**

4. 3. Trend comparisons

Objective: **study the effect of increasing levels of a factor**

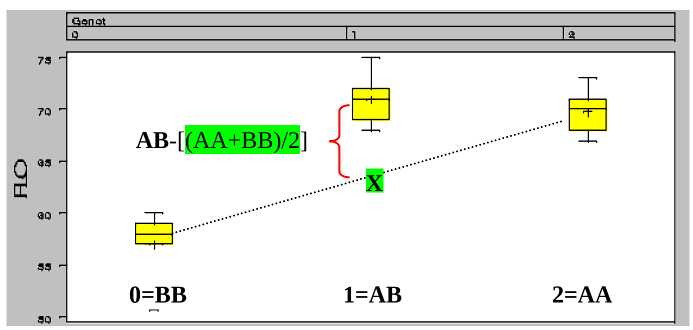
The experimenter is interested in the dose response relationship.

The statistical analysis should **not** be concerned with **pairwise comparisons**.

Examples for orthogonal contrasts. Genetic examples. (Table 4.6. notes)

- **0** dose of allele A (homozygous **BB** individuals)
- **1** dose of allele A (heterozygous **AB** individuals)
- **2** doses of allele A (homozygous **AA** individuals)

```
Data contrast;
Input Genot Nitrogen Flowering;
Cards;
0     12.0       58
0     12.5       51
0     12.1       57
0     11.8       59
0     12.6       60
1     13.5       71
1     13.8       75
1     13.0       69
1     13.2       72
1     13.0       68
1     12.8       73
1     12.9       69
1     13.4       70
1     12.7       71
1     13.6       72
2     13.8       73
2     14.5       68
2     13.9       70
2     14.2       71
2     14.1       67
;
proc glm;
  class genot;
  model Nitrogen Flowering= genot;
  contrast 'Lineal'    genot -1  0  1;
  contrast 'Quadratic' genot  1 -2  1;
run; quit;
```

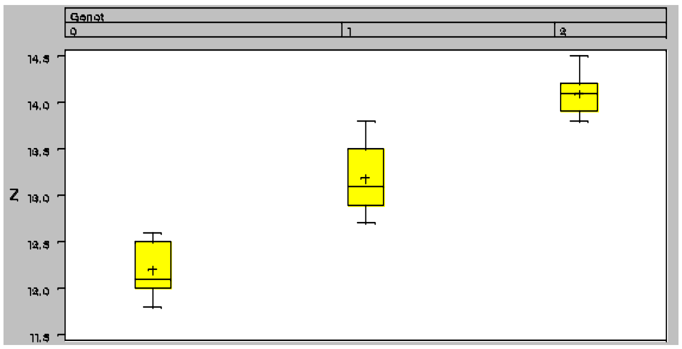


Lineal contrast: BB vs. AA
Quadratic contrast: AB vs. average of AA + BB

ANOVA dependent variable: Nitrogen

Source	DF	SS	MS	F Value	Pr > F
Model	2	9.033	4.5165	38.60	0.0001
Error	17	1.989	0.1170		
Corrected Total	19	11.022			

Contrast	DF	Contrast SS	MS	F Value	Pr > F
Lineal	1	**9.025**	9.0250	**77.14**	0.0001
Quadratic	1	**0.008**	0.0080	0.07	0.7969



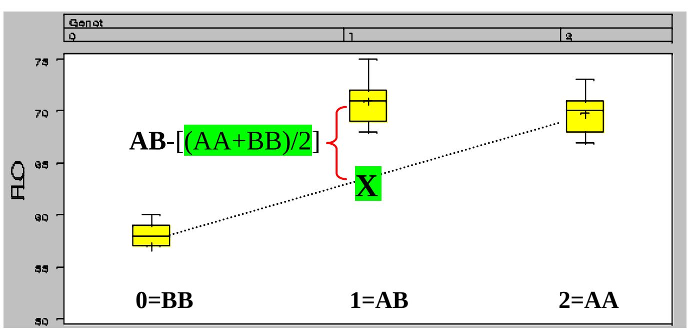
$$\frac{\text{SS lineal} + \text{SS quadratic.}}{2} = \textbf{MS Model}$$

The ANOVA **MS Model** is the **average** of the two effects.

ANOVA dependent variable: Flowering

Source	DF	SS	MS	F Value	Pr > F
Model	2	698.4	349.2	52.63	0.0001
Error	17	**112.8**	6.6		
Corrected Total	19	811.2			

Contrast	DF	Contrast SS	MS	F Value	Pr > F
Lineal	1	**409.6**	409.6	**61.73**	0.0001
Quadratic	1	**288.8**	288.8	**43.52**	0.0001



Regression analysis

Source	DF	SS	MS	F Value	Pr > F
Model	1	**409.6**	409.6	**18.36**	0.0004
Error	18	**401.6**	22.3		
Total	19	811.2			

401.6= 112.8 + **288.8**

Example

ST&D Table 15.11 Page 387

Yield of Ottawa Mandarin soybeans grown in MN, in bushels per acre.

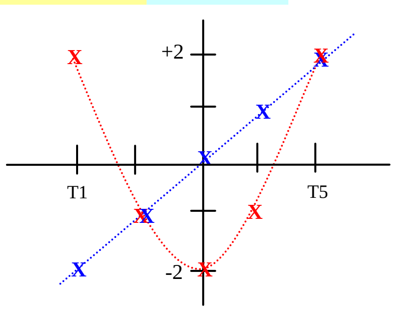
	Row spacing (in inches)				
Rep.*	**18**	**24**	**30**	**36**	**42**
1	**33.6**	**31.1**	**33.0**	**28.4**	**31.4**
2	**37.1**	**34.5**	**29.5**	**29.9**	**28.3**
3	**34.1**	**30.5**	**29.2**	**31.6**	**28.9**
4	**34.6**	**32.7**	**30.7**	**32.3**	**28.6**
5	**35.4**	**30.7**	**30.7**	**28.1**	**29.6**
6	**36.1**	**30.3**	**27.9**	**26.9**	**33.4**
Means	**31.15**	**31.63**	**30.17**	**29.53**	**30.03**

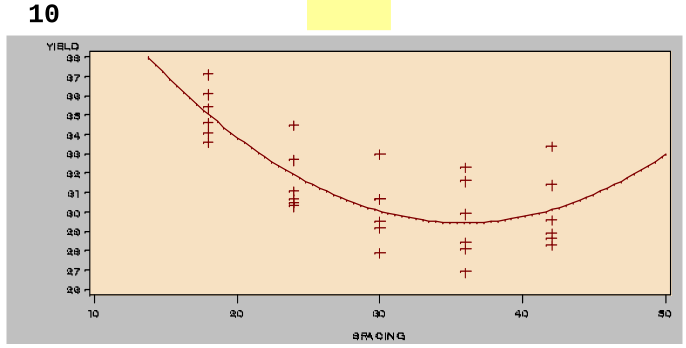
* Original example with blocks, treated as reps in this example

Coefficients for trend comparisons for **equally spaced** treatments

ST&D p390 and PLS205 WEB page Lecture

No. of treatments	Degree polynom.	T1	T2	T3	T4	T5	T6	Σc_i^2	$(\sum c_i \bar{Y}_{i.})^2$	$\frac{(\sum c_i \bar{Y}_{i.})^2}{(\sum c_i^2)/r}$
2	**1**	-1	+1					2		
3	**1**	-1	0	+1				2		
	2	+1	-2	+1				6		
4	**1**	-3	-1	+1	+3			20		
	2	+1	-1	-1	+1			4		
	3	-1	+3	-3	+1			20		
5	**1**	-2	-1	0	+1	+2		**10**	152.3	**91.3 ****
	2	+2	-1	-2	-1	+2		**14**	78.5	**33.7 ****
	3	-1	+2	0	-2	+1		**10**	0.8	**0.5 NS**
	4	+1	-4	+6	-4	+1		**70**	2.4	**0.2 NS**
6	**1**	-5	-3	-1	+1	+3	+5	70		
	2	+5	-1	-4	-4	-1	+5	84		
	3	-5	+7	+4	-4	-7	+5	180		
	4	+1	-3	+2	+2	-3	+1	28		
	5	-1	+5	-10	+10	-5	+1	252		





Unequally spaced treatments using multiple regression

```
data stp387reg;
title 'Multiple regression CRD';
input S yield;
cards;
18 33.6  …;
proc glm;
  model yield= S S*S S*S*S S*S*S*S;
run; quit;
```

(note the absence of a class statement in the regression analysis)

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	125.7	31.4	9.90	<.0001
Error	25	79.3	3.2		
Corrected Total	29	205.0			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
S	1	91.3	91.3	28.76	<.0001
S*S	1	33.7	33.7	10.62	0.0032
S*S*S	1	0.5	0.5	0.16	0.6936
S*S*S*S	1	0.2	0.2	0.06	0.8052

Same as

```
data stp387reg;
title 'Contrast CRD';
input  S yield;
      cards;
18 33.6 …;
proc glm;
 class S;
 model yield=S;
  contrast 'linear'      S -2 -1  0 +1 +2;
  contrast 'Quadratic'   S +2 -1 -2 -1 +2;
  contrast 'Cubic'       S -1 +2  0 -2 +1;
  contrast 'Quartic'     S +1 -4 +6 -4 +1;
run; quit;
```

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	125.7	31.4	9.90	<.0001
Error	25	79.3	3.2		
Corrected Total	29	205.0			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
S	1	91.3	91.3	28.76	<.0001
S*S	1	33.7	33.7	10.62	0.0032
S*S*S	1	0.5	0.5	0.16	0.6936
S*S*S*S	1	0.2	0.2	0.06	0.8052

Same result in both analyses. The multiple regression analysis can be used with unequally spaced treatments, but the Contrast analysis not.