ANALYSIS OF COVARIANCE

Experimental error is due to variability among experimental units. To increase precision, minimize experimental error.

Plan A: Blocking

- Maximize differences among blocks and minimize differences within blocks
- Does not work if heterogeneity does not follow a pattern, eg variation in plant stand, spotty soil heterogeneity
- Blocking isolates variation known before the experiment is installed. Does not work if variability occurs after blocking, eg insect or disease damage

Plan B: Covariance analysis

- Covariance isolates variation that occurs after the experiment is installed or that could not be isolated by blocking.
- Corrects for variability in observations Y related to measured variability in factor X - the covariate
- Combines analysis of variance and regression analysis
- Y can be adjusted linearly based on size of X for that observation
- Can adjust for sources of bias in observational studies

Covariance is not interchangeable with blocking.
Covariance can be combined with blocking.

Covariance

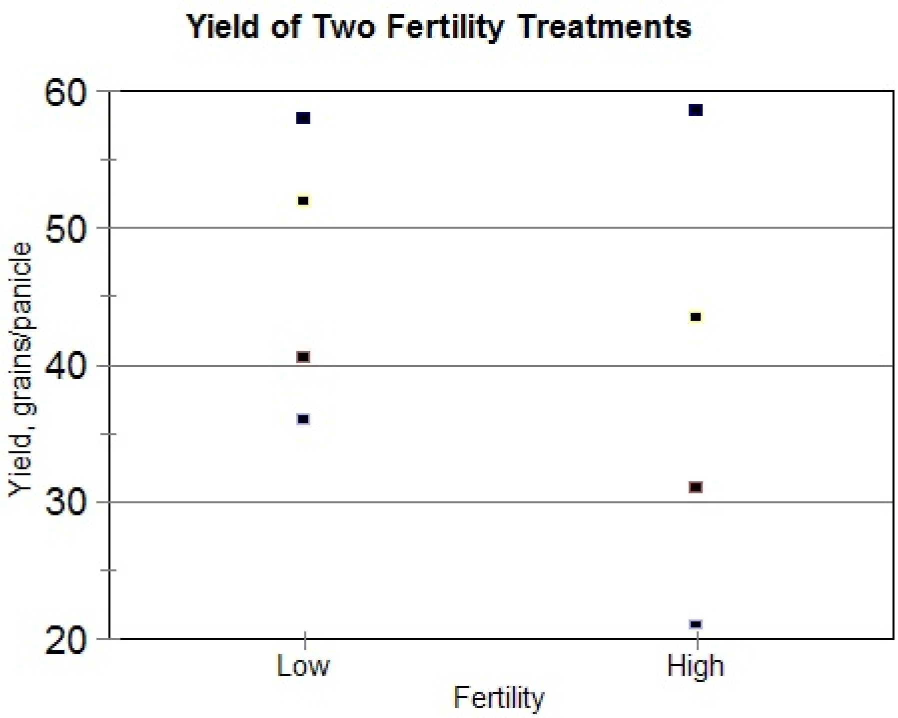
Measurements are made on the observed response Y and also on an additional variable X, called the covariate.

Covariate, X

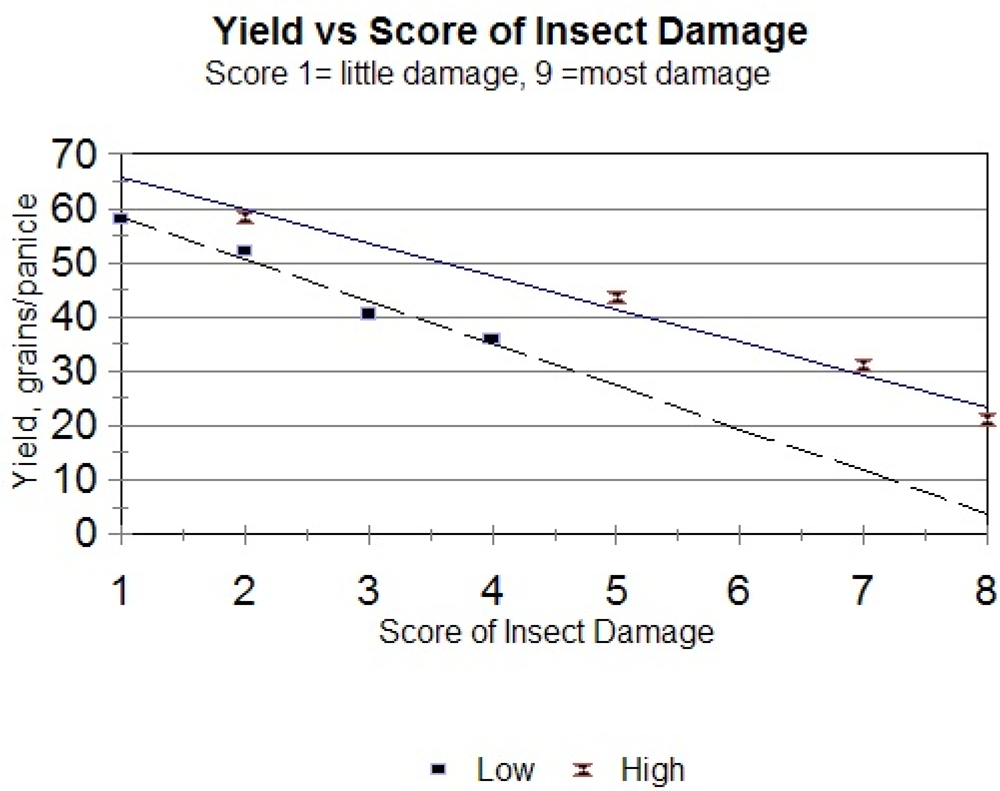
- Must be quantitative
- Measures differences among experimental units
- Can be measured before or during experiment
- Assumed to be linearly related to Y ($Y = a + bX$)

Analysis of covariance adjusts the values of Y for variation in the covariate X.

Example: An experiment measuring the yield of rice in high and low fertility treatments was affected by insect damage. Initial analysis showed a mean of 46.3 for the low fertility treatment and a mean of 38.4 for the high fertility treatment. Regression was done to determine the effect of insect damage on yield. After correcting for the effect of insect damage on yield, the high fertility treatment was shown to produce higher yields.



Yield of Two Fertility Treatments
60
50
40
30
20
Yield, grains/panicle
Low
High
Fertility



Yield vs Score of Insect Damage
Score 1= little damage, 9 =most damage
70
60
50
40
30
20
10
0
Yield, grains/panicle
1
2
3
4
5
6
7
8
Score of Insect Damage
Low
High

Relationship of Covariate to Treatment

Covariate can be independent of treatment, eg initial weight of animals, insect infestation

Adjust for effects of covariate to reduce error

Covariate can be affected by treatment

Use to investigate mechanism or nature of treatment effects

eg water management of rice

depth of water affects both grain yield and weed population

is effect on grain yield primarily due to effect on weed population?

is there a significant difference in yield after correcting for weed population?

eg how much of the increase in crop yield following manure application is due to nematode control?

eg physiological mechanisms: can the effect of growth hormone be explained by the increase in IGF-1?

Covariance should be used whenever there is an important independent variable that can be measured but can not be controlled by blocking. For example:

Plant stand
Severity of pest damage
Initial size of plants
Stage of development
Soil properties
Initial weight of animals
Food intake
Environmental effects

Uses of Covariance:

1. To assist in the interpretation of data.

2. To control error and increase precision.

3. To adjust treatment means of the dependent variable for differences in values of the corresponding independent variable.

4. To partition a total covariance or sum of cross products into component parts.

5. To estimate missing data.

Mathematical Model

The mathematical model for covariance is that for the analysis of variance plus an additional term for the independent variable.

For the Randomized Complete Block Design

$$Y_{ij} = \overline{Y}.. + T_i + B_j + b(X_{ij} - \overline{X}..) + e_{ij}$$

Where Y_{ii} is the observed dependent variable

- $\overline{Y}..$ is the experiment mean of the dependent variable
- T_i is the treatment effect
- B_j is the block effect
- b is the regression coefficient of the relationship between the dependent variable, Y and the independent variable, X
- X_{ii} is the observed independent variable
- $\overline{X}..$ is the experiment mean of the independent variable
- e_{ij} is the residual variance or experimental error

Assumptions in Covariance

1. The X's are fixed and measured without error. (Not always true)

2. The regression of Y on X after removal of block and treatment differences is linear and independent of treatments and blocks.

3. The residuals are normally and independently distributed with zero mean and a common variance.

Steps in Covariance Analysis - RCBD

1. Construct ANOVA tables as RCBD for X, independent variable or covariate, and for Y, dependent variable

2. Check for treatment effect on X and on Y using F-test

3. Calculate sums of cross-products

4. Construct Analysis of Covariance table including sums of squares for X and Y, and sums of cross-products. Include Trt+Err df, SSX, SP and SSY

5. Calculate SSRegr (adj for trt) and SSDev(Regr+Trt)

6. Calculate SSRegr (trt + err) and SSTrt (adj for regr)

7. Complete the Analysis of Covariance table and test MSRegr (adj for trt) and MSTrt (adj

for regr) against MSDevRegr (the remaining error)

8. Adjust treatment means

Example Problem - Covariance in a RCBD

Initial Weights, X, and Kidney Fat, Y, for 16 steers given 4 hormone treatments in an RCBD.

	Hormone							
	1		2		3		4	
Block	X	Y	X	Y	X	Y	X	Y
1	56	133	44	128	53	129	69	134
2	47	132	44	127	51	130	42	125
3	41	127	36	127	38	124	43	126
4	50	132	46	128	50	129	54	131
Total	194	524	170	510	192	512	208	516

ANOVA for X

	Hormone				
Block	1	2	3	4	Total
1	56	44	53	69	222
2	47	44	51	42	184
3	41	36	38	43	158
4	50	46	50	54	200
Total	194	170	192	208	764

ANOVA Table for X

Source	df	SSX	MS	F	$F_{.05}$
Total	15	973			
Block	3	545	181.67	6.73	3.86
Hormone	3	185	61.67	2.28	3.86
Error	9	243	27.00		

ANOVA for Y

Block	Hormone 1	2	3	4	Total
1	133	128	129	134	524
2	132	127	130	125	514
3	127	127	124	126	504
4	132	128	129	131	520
Total	524	510	512	516	2062

ANOVA Table for Y

Source	df	SS	MS	F	$F_{.05}$
Total	15	127.75			
Block	3	56.75	18.92	4.03	3.86
Hormone	3	28.75	9.58	2.04	3.86
Error	9	42.25	4.69		

Calculation of Sums of Cross-Products

$$\text{Correction factor, } cf = \frac{\Sigma X \Sigma Y}{n} = \frac{(764)(2062)}{16} = 98460.5$$

$$SPTotal = \Sigma XY - cf = (56)(133) + \ldots + (54)(131) - 98460.5 = 295.5$$

$$SPBlock = \frac{\Sigma X_{.j} \Sigma Y_{.j}}{r} - cf = \frac{(222)(524) + (200)(520)}{4} - 98460.5 = 173.5$$

$$SPHormone = \frac{\Sigma X_{i.} \Sigma Y_{i.}}{r} - cf = \frac{(194)(524) + (208)(516)}{4} - 98460.5 = 36.5$$

$$SPError = SPTotal - SPHormone = 295.5 - 173.5 - 36.5) = 85.5$$

ANALYSIS OF COVARIANCE

Sums of Squares and Products

Source	df	SSX	SP	SSY
Total	15	973	295.5	127.75
Block	3	545	173.5	56.75
Hormone	3	185	36.5	28.75
Error	9	243	85.5	42.25
Hormone + Error	12	428	122.0	71.00

$$SSRegr = r^2 * SSY = \frac{SP^2}{SSX}$$

For Y, SSTrt+Err = 71, with 12 df. This can be divided in 2 ways:
1) SSTrt + SSRegr (adj for trt) + SSDev = 71
2) SSRegr + SSTrt (adj for regr) + SSDev = 71

1. First adjust for Trt, then subdivide Error into Regr (adj for trt) and Dev from Regr and Trt

$$SSTrt = 28.75$$

$$SSError(adj\ Trt) = 42.25$$

$$SSRegr(adj\ Trt) = \frac{SPError^2}{SSXError} = \frac{85.5^2}{243} = 30.08$$

$$SSDev = SSError(adj\ Trt) - SSRegr(adj\ Trt) = 42.25 - 30.08 = 12.17$$

2. First adjust for Regr, then subdivide Error into Trt (adj for Regr) and Dev from Regr and Trt

$$SSRegr = \frac{SPTrt{+}Error^2}{SSXTrt{+}Error} = \frac{122^2}{428} = 34.78$$

$$SSTrt(adj\ Regr) = SSTrt{+}Error - SSRegr - SSDev$$

$$= 71.00 - 34.78 - 12.17 = 24.05$$

Note that Trt and Regr are not orthogonal, so the SS depends on which is calculated first. F-tests are conducted on the adjusted values.

Completion of ANCOVA

Source	df	SS	MS	F	$F_{.05}$	$F_{.01}$
Trt+Err	(12)	71				
Trt	3	28.75				
Regr (adj Trt)	1	30.08	30.08	19.79	5.32	11.26
Dev	8	12.17	1.52			
Regr	1	34.78				
Trt (adj Regr)	3	24.05	8.02	5.27	4.06	7.59
Dev	8	12.17	1.52			

Adjustment of Treatment Means

$$\text{Adjusted Mean, } \overline{Y}_{i.}(\text{adj}) = \overline{Y}_{i.} - b(\overline{X}_{i.} - \overline{X}_{..})$$

$$b = \frac{\text{SPErr}}{\text{SSXErr}} = \frac{85.5}{243} = 0.352$$

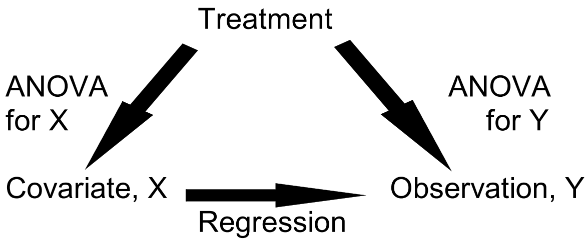
Diets	$\overline{Y}_{i.}$	$\overline{X}_{i.}$	$\overline{X}_{i.} - \overline{X}_{..}$	$b(\overline{X}_{i.} - \overline{X}_{..})$	$\overline{Y}_{i.}(\text{adj})$
1	131.0	48.5	0.75	0.26	130.7
2	127.5	42.5	-5.25	-1.85	129.3
3	128.0	48.0	0.25	0.088	127.9
4	129.0	52.0	4.25	1.50	127.5

Summary

Adjusting for covariate (initial weight)

- Reduced MSError, increasing precision
- Increased MSTreatment, increasing power
- Compensated for the fact that treatment 2 steers had lower mean initial weight and treatment 4 steers had a higher initial weight, chance occurrences which tended to obscure hormone effects.

Interpretation of Covariance Analysis



If the regression of Y on X is not significant (or if $P > .1$), remove the covariate from the model. If the regression is significant, examine the effects of treatment on X and Y.

Treatment effects on:

X	Y		Conclusions
	Before Covariance	After Covariance	
ns	sig	ns	Apparent treatment effect due to variation in X.
ns	ns	sig	True treatment effects obscured by variation in X.
sig	sig	ns	Apparent treatment effect may be due to a treatment effect on X, which then affects Y (the mechanism of the treatment action may be via X).
sig	sig	sig	Treatment had significant effect on Y beyond that due to variation in X.

Using Covariance for Calculating Missing Data
(Adapted from Steel and Torrie)

Gives an unbiased estimate of treatment and error sum of squares
Leads to an unbiased test of treatment means
Analysis is convenient and simple

Mean ascorbic acid content of three 2 g samples of turnip greens in mg/100 g dry weight

	Block (Day)										
Trt	1		2		3		4		5		Total
	Y	X	Y	X	Y	X	Y	X	Y	X	
A	0	1	887	0	897	0	850	0	975	0	3609
B	857	0	1189	0	918	0	968	0	909	0	4841
C	917	0	1072	0	975	0	930	0	954	0	4848
Totals	1774	1	3148	0	2790	0	2748	0	2838	0	13298

Procedure:

1. Set Y = 0 for the missing plot
2. Define covariate as X = 0 for an observed Y, and X = +1 (or -1) for Y = 0
3. Carry out the analysis of covariance to obtain the error sums of squares and products
4. Compute B = SP/SSX and change sign to estimate the missing value.

ANCOVA

Source	df	SSX	SP	SSY
Total	14	0.9333	-886.53	945.296
Block	4	0.2667	-295.20	359.823
Treatment	2	0.1333	-174.73	203.533
Error	8	0.5333	-426.60	381.940

b = SP/SSX = (-426.60/0.5333) = -799.92 = -800
Change the sign -800 (-1) = 800

800 is the missing value.

Redo the ANOVA with this value. Subtract 1 df from error value and total df.
Can calculate several missing data by introducing a new independent variable for each missing datum and using multiple covariance.