

The General Linear Model consists of

- a response variable,
- one or more explanatory variables,
- parameters that relate response to explanatory variable(s),
- an error term.

1. Steel and Torrie (Principles and Procedures of Statistics 1960) report oat yield (Yield = bushels/acre) of untreated seeds compared to seeds treated with Panogen (tr = Panogen or not). Both uninfected seeds and seeds infected with H. victoriae were used (inf = infected or not).

Y = response
X = Tr, Inf

1a. Assign symbols to the response and explanatory variables.

1b. Write a General linear model for the analysis.

$$y = \beta_0 + \beta_{inf} \cdot Inf + \beta_{Tr} \cdot Tr + \beta_{Inf \cdot Tr} \cdot Inf \cdot Tr + error$$

1c. Fill in the ANOVA table below.

```
MTB > anova 'yield' = 'inf' 'tr' 'inf'*'tr';
```

Factor	Type	Levels	Values
inf	fixed	2	0 1
tr	fixed	2	0 1

Analysis of Variance for yield

Source	DF	SS	MS	F	P
inf	1	486.20	486.20	4.36	0.059
tr	1	145.20	145.20	1.30	0.276
inf*tr	1	57.00	57.00	0.511	0.488
Error	12	1338.75	111.56		
Total	15	2027.16			

1d. What is the variance in Yield ?

$$\text{var(Yield)} = 2027.16 / 15 = 135.14$$

2. Draw an example of an unacceptable plot of residual versus fitted values.

Vertical dispersion of residuals not uniform across plot.
Cone shaped dispersion is typical