

STAT22200 Spring 2014 Chapter 09A

Yibi Huang

May 20, 2014

- 9.1 Contrasts for Factorial Data
- 9.2.2 One-Cell Interaction
- 9.2.3 Quantitative Factors

9.1 Contrasts for Factorial Data

Recall in a one-way model, a *contrast* is a linear combination of treatment mean μ_i 's

$$C = \sum_{i=1}^g w_i \mu_i \quad \text{such that} \quad \sum_{i=1}^g w_i = 0$$

Similar in factorial data, say 3-way data, a *contrast* is a linear combination of treatment. μ_{ijk} 's

$$C = \sum_{ijk} w_{ijk} \mu_{ijk} \quad \text{such that} \quad \sum_{ijk} w_{ijk} = 0$$

For example, for 3-way data, a natural estimator for the contrast $C = \sum_{ijk} w_{ijk} \mu_{ijk}$ is

$$\hat{C} = \sum_{ijk} w_{ijk} \hat{\mu}_{ijk} = \sum_{ijk} w_{ijk} \bar{y}_{ijk\bullet}$$

By the independence of $\bar{y}_{ijk\bullet}$'s, we know

$$\begin{aligned} \text{Var}\left(\sum_{ijk} w_{ijk} \bar{y}_{ijk\bullet}\right) &= \sum_{ijk} \text{Var}(w_{ijk} \bar{y}_{ijk\bullet}) \\ &= \sum_{ijk} w_{ijk}^2 \text{Var}(\bar{y}_{ijk\bullet}) = \sum_{ijk} w_{ijk}^2 \frac{\sigma^2}{n_{ijk}}. \end{aligned}$$

Here n_{ijk} is the number of replicates in treatment (i, j, k) . The estimator \hat{C} has standard deviation and standard error

$$\text{SD}(\hat{C}) = \sigma \sqrt{\sum_{ijk} \frac{w_{ijk}^2}{n_{ijk}}}, \quad \text{SE}(\hat{C}) = \sqrt{\text{MSE}} \sqrt{\sum_{ijk} \frac{w_{ijk}^2}{n_{ijk}}}$$

A $(1 - \alpha)100\%$ confidence interval for the contrast C is

$$\hat{C} \pm t_{dfE, \alpha/2} \times \text{SE}(\hat{C})$$

where dfE represents the degrees of freedom for SSE.

Commonly Used Contrasts

- *simple effect*: comparing two treatments that have the same levels of all factors but one, e.g., $C = \mu_{i_1jk} - \mu_{i_2jk}$
- *main-effect contrast*: when the coefficients w_{ijk} depend only on a single index. For example, say $w_{ijk} = c_i$ depend only on index i for all i, j, k . Then

$$C = \sum_{ijk} w_{ijk} \mu_{ijk} = \sum_{ijk} c_i \mu_{ijk} = \sum_i c_i \sum_{jk} \mu_{ijk} = \sum_i c_i \mu_{i\bullet\bullet}$$

Recall $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$.

Summing over indexes i and j , all other terms vanish except μ and α_i because of the zero-sum constraints,

$$\begin{aligned} \mu_{i\bullet\bullet} &= bc\mu + bc\alpha_i + c\beta_{\bullet} + b\gamma_{\bullet} + c\alpha\beta_{i\bullet} + b\alpha\gamma_{i\bullet} + \beta\gamma_{\bullet\bullet} + \alpha\beta\gamma_{i\bullet\bullet} \\ &= bc(\mu + \alpha_i). \end{aligned}$$

Thus such a contrast only depends on main effects for factor A.

$$C = \sum_i c_i \mu_{i\bullet\bullet} = bc \underbrace{\sum_i c_i \mu}_{=0} + bc \sum_i c_i \alpha_i = bc \sum_i c_i \alpha_i$$

- *interaction contrasts contrast*: the coefficients depend only on the indices of factors in the interaction in question and satisfy the same zero-sum restrictions as their corresponding model terms.

E.g., the contrast $C = \sum_{ijk} w_{ijk} \mu_{ijk}$ with $w_{ijk} = c_{ij}$ depending only on indexes i and j and satisfies $\sum_i c_{ij} = \sum_j c_{ij} = 0$ for all i, j is a two-way interaction contrast.

$$C = \sum_{ijk} w_{ijk} \mu_{ijk} = \sum_{ijk} c_{ij} \mu_{ijk} = \sum_{ij} c_{ij} \sum_k \mu_{ijk} = \sum_{ij} c_{ij} \mu_{ij\bullet}$$

Summing μ_{ijk} over index k , all terms involve index k vanish because of the zero-sum constraints,

$$\begin{aligned} \mu_{ij\bullet} &= c\mu + c\alpha_i + c\beta_j + \gamma_{\bullet} + c\alpha\beta_{ij} + \alpha\gamma_{i\bullet} + \beta\gamma_{j\bullet} + \alpha\beta\gamma_{ij\bullet} \\ &= c(\mu + \alpha_i + \beta_j + \alpha\beta_{ij}). \end{aligned}$$

Thus such a contrast compares the AB interactions.

$$\begin{aligned} C &= \sum_{ij} c_{ij} \mu_{ij\bullet} = c \sum_{ij} c_{ij} (\mu + \alpha_i + \beta_j + \alpha\beta_{ij}) \\ &= c(c_{\bullet\bullet}\mu + \sum_i c_{i\bullet}\alpha_i + \sum_j c_{\bullet j}\beta_j + \sum_{ij} c_{ij}\alpha\beta_{ij}) = c \sum_{ij} c_{ij}\alpha\beta_{ij}. \end{aligned}$$

9.2.2 One-Cell Interaction

Example 9.2, p.211 Oehlert

Table 9.2: Data from a replicated four-factor experiment. All factors have two levels, labeled low and high.

A	B	C	D			
			Low		High	
low	low	low	26.1	27.5	23.5	21.1
low	low	high	22.8	23.8	30.6	32.5
low	high	low	22.0	20.2	28.1	29.9
low	high	high	30.0	29.3	38.3	38.5
high	low	low	11.4	11.0	20.4	22.0
high	low	high	22.3	20.2	28.7	28.8
high	high	low	18.9	16.4	26.6	26.5
high	high	high	29.6	29.8	34.5	34.9

Data:

<http://users.stat.umn.edu/~gary/book/fcdae.data/exmpl9.2>

```

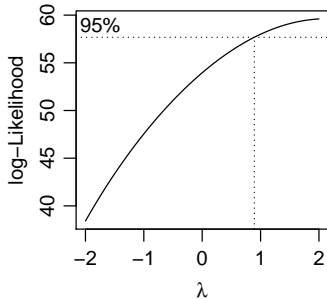
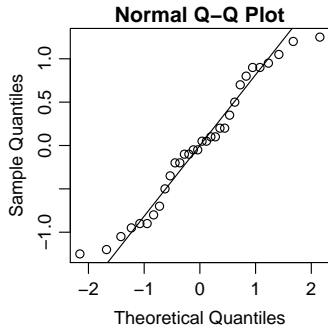
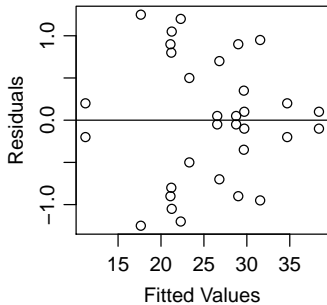
> mydata =
  read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/exmpl9.2",h=T)
> attach(mydata)
> lm1 = lm(y ~ A*B*C*D)
> anova(lm1)

```

Analysis of Variance Table

Response: y

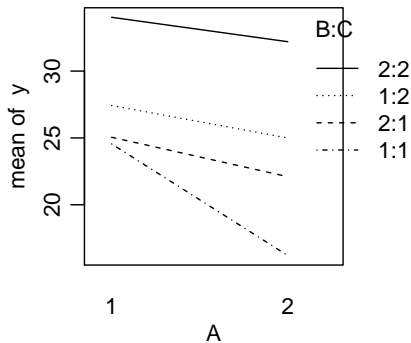
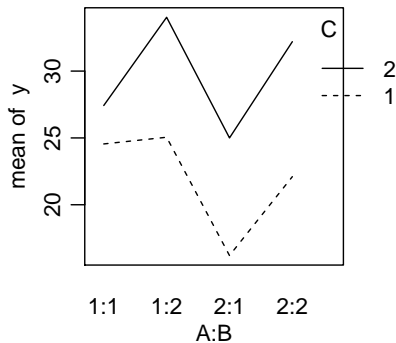
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	120.90	120.90	117.4511	8.871e-09	***
B	1	204.02	204.02	198.1979	1.970e-10	***
C	1	472.78	472.78	459.2896	3.288e-13	***
D	1	335.40	335.40	325.8336	4.621e-12	***
A:B	1	18.00	18.00	17.4863	0.0007050	***
A:C	1	24.85	24.85	24.1421	0.0001559	***
B:C	1	27.38	27.38	26.5987	9.541e-05	***
A:D	1	15.12	15.12	14.6934	0.0014664	**
B:D	1	10.81	10.81	10.5027	0.0051192	**
C:D	1	6.48	6.48	6.2951	0.0232492	*
A:B:C	1	11.52	11.52	11.1913	0.0041075	**
A:B:D	1	34.03	34.03	33.0601	2.985e-05	***
A:C:D	1	50.00	50.00	48.5732	3.161e-06	***
B:C:D	1	22.11	22.11	21.4803	0.0002754	***
A:B:C:D	1	13.78	13.78	13.3880	0.0021183	**
Residuals	16	16.47	1.03			



The diagnostic plots look fine.

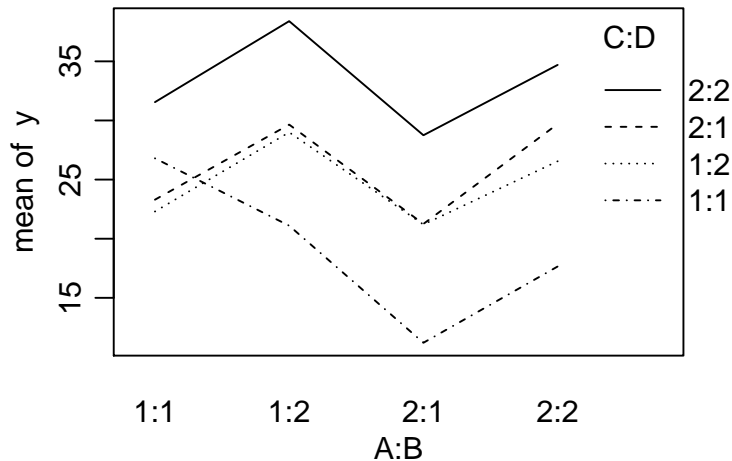
How to Visualize 3-Way Interactions?

```
> A = as.factor(A)
> B = as.factor(B)
> C = as.factor(C)
> D = as.factor(D)
> interaction.plot(A:B, C, y)
> interaction.plot(A, B:C, y)
```



How to Visualize 4-Way Interactions?

```
> interaction.plot(A:B, C:D, y)
```



Except the treatment with $A = B = C = D = 1$, the effect for A, B, C, and D seems to be additive.

Consider the model

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta\gamma\delta)_{1111} + \varepsilon_{ijklm}$$

```
> lmonecell = lm(y ~ A+B+C+D+I(A==1&B==1&C==1&D==1))
```

```
> anova(lmonecell)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	120.90	120.90	94.653	3.745e-10	***
B	1	204.02	204.02	159.726	1.318e-12	***
C	1	472.78	472.78	370.136	< 2.2e-16	***
D	1	335.40	335.40	262.586	4.172e-15	***
I(A == 1 & B == 1 & C == 1 & D == 1)	1	217.35	217.35	170.162	6.449e-13	***
Residuals	26	33.21	1.28			

```
> anova(lmonecell,lm1)
```

Analysis of Variance Table

```
Model 1: y ~ A + B + C + D + I(A == 1 & B == 1 & C == 1 & D == 1)
```

```
Model 2: y ~ A * B * C * D
```

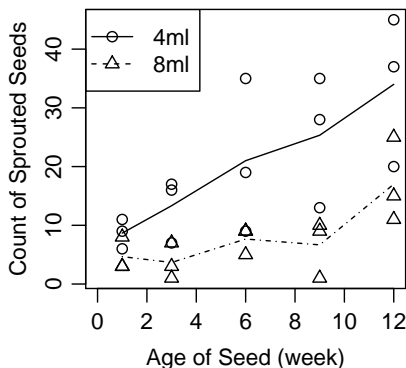
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	26	33.21				
2	16	16.47	10	16.74	1.6263	0.1861

9.2.3 Quantitative Factors

Sprouting Barley Example Revisit (p.166, Oehlert)

Recall the sprouting barley study on the conditions barley germinate. The response is the number of seeds germinating in 100 seeds.

water	Age of Seeds (weeks)				
	1	3	6	9	12
4(ml)	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
8(ml)	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25



Sprouting Barley Example Revisit (2)

```
> barley = read.table("SproutingBarley.txt",header=T)
> lmfull = lm(y ~ as.factor(week)*as.factor(water),data=barley)
> anova(lmfull)
```

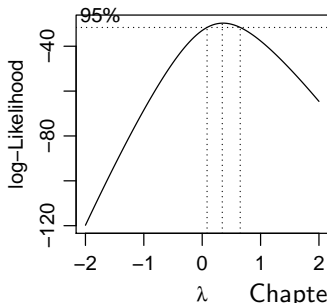
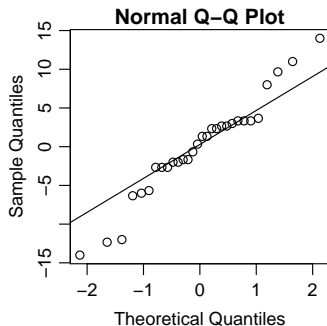
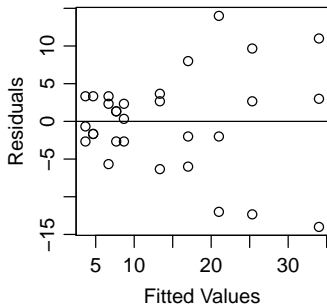
Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(week)	4	1321.13	330.28	5.5293	0.003645	**
as.factor(water)	1	1178.13	1178.13	19.7232	0.000251	***
as.factor(week):as.factor(water)	4	208.87	52.22	0.8742	0.496726	
Residuals	20	1194.67	59.73			

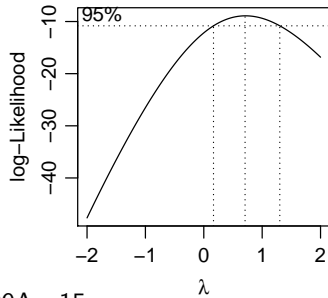
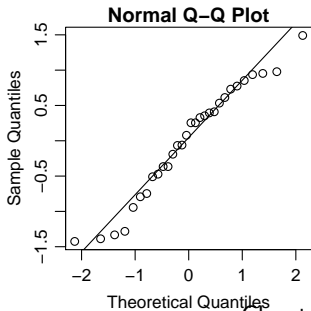
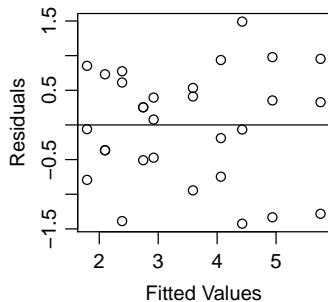
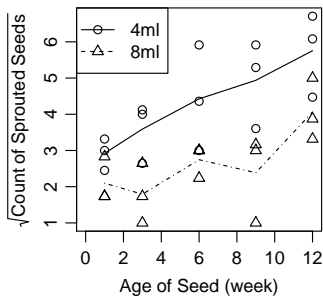
Before making conclusions, let's check model assumptions.

Sprouting Barley Example Revisit (3)



Spot any problem?
Remedy?

Sprouting Barley Example Revisit (4)



Sprouting Barley Example Revisit (4)

```
> lmfull2 = lm(sqrt(y) ~ as.factor(week)*as.factor(water), data=barley)
```

```
> anova(lmfull2)
```

Analysis of Variance Table

Response: sqrt(y)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(week)	4	21.8949	5.4737	5.9406	0.002555	**
as.factor(water)	1	21.8930	21.8930	23.7606	9.177e-05	***
as.factor(week):as.factor(water)	4	2.2485	0.5621	0.6101	0.660139	
Residuals	20	18.4280	0.9214			

- ▶ Now what is your conclusion?
- ▶ The main effect “seed age” being significant just means that seeds of different ages (1, 3, 6, 9, and 12 weeks) have *different* germination rates. It doesn't even tell us whether sprouting rate *increases* with the age of seeds.
- ▶ ANOVA models treat all factors as *categorical*.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

- ▶ Are there models that treat quantitative factors as quantitative?

9.2.3 Quantitative Factors

Recall in Lecture 6, when treatments in a CRD are quantitative, we can fit a polynomial model

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_{ij}.$$

In a two-way $a \times b$ factorial design, if factor A is quantitative, with a quantitative levels at x_1, \dots, x_a , and factor B is categorical, we may consider a polynomial model

$$y_{ijk} = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_m x_i^m + \beta_j + \varepsilon_{ijk}.$$

If

$$\mu + \alpha_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_m x_i^m \quad \text{for all } i = 1, \dots, a,$$

then the polynomial model is equivalent to the additive model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}.$$

- ▶ The order of the polynomial m cannot exceed $a - 1$. (Why?)
- ▶ As long as $m \leq a - 1$, the polynomial model is *nested* in the additive model $y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$. (Why?)

If Both Factors are Quantitative . . .

In a two-way $a \times b$ factorial design, if the two factors are both quantitative, say factor A has a levels at x_1, \dots, x_a , and factor B has b levels at z_1, \dots, z_b , we may consider a polynomial model

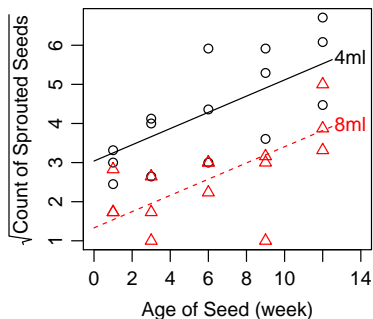
$$y_{ijk} = \theta_0 + \theta_1 x_i + \dots + \theta_m x_i^m + \phi_1 z_j + \dots + \phi_r z_j^r + \varepsilon_{ijk}.$$

- ▶ the order m of the polynomial for factor A cannot exceed $a - 1$,
- ▶ the order r of the polynomial for factor B cannot exceed $b - 1$
- ▶ this is an additive model
- ▶ As long as the orders of polynomial $m \leq a - 1$ and $r \leq b - 1$, the polynomial model is *nested* in the additive model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}. \text{ (Why?)}$$

(Of course, it is also nested in the main-effect-interaction model $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$.)

Polynomial Model for the Sprouting Barley Example (1)



Both factors “week” and “water” are quantitative. The ANOVA table shows no significant interaction. From the scatter plot, we see the square-root transformed response is roughly linearly with the age of seeds (week). We thus fit the model

$$y_{ijk} = \theta_0 + \theta_1 \text{week}_i + \phi_1 \text{water}_j + \varepsilon_{ijk}.$$

```
> lmadd1 = lm(sqrt(y) ~ week + water, data=barley); lmadd1
```

Coefficients:

(Intercept)	week	water
4.7481	0.2074	-0.4271

The fitted model is

$$\hat{y}_{ijk} = 4.7481 + 0.2074 \text{week}_i - 0.4271 \text{water}_j.$$

Polynomial Model for the Sprouting Barley Example (2)

Compare the ANOVA tables of the 3 models.

Full Model: $\sqrt{y_{ijk}} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(week)	4	21.8949	5.4737	5.9406	0.002555	**
as.factor(water)	1	21.8930	21.8930	23.7606	9.177e-05	***
as.factor(week):as.factor(water)	4	2.2485	0.5621	0.6101	0.660139	
Residuals	20	18.4280	0.9214			

Additive Model: $\sqrt{y_{ijk}} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(week)	4	21.895	5.4737	6.3536	0.001236	**
as.factor(water)	1	21.893	21.8930	25.4121	3.746e-05	***
Residuals	24	20.677	0.8615			

Polynomial Model: $\sqrt{y_{ijk}} = \theta_0 + \theta_1 \text{week}_i + \phi_1 \text{water}_j + \varepsilon_{ijk}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
week	1	20.346	20.3464	24.718	3.286e-05	***
water	1	21.893	21.8930	26.597	1.997e-05	***
Residuals	27	22.225	0.8231			

Compare the SSE and MSE of the 3 models. Which one has the minimal SSE? Which one has the minimal MSE?

Polynomial Model for the Sprouting Barley Example (3)

Do we need to consider an higher order polynomial model?

Since the polynomial model is nested in the full effects model, we can perform a goodness-of-fit test

$$F = \frac{(SSE_{reduced} - SSE_{full}) / (df_{reduced} - df_{full})}{SSE_{full} / df_{full}}$$

```
> lmadd1 = lm(sqrt(y) ~ week + water, data=barley)
> lmfull2 = lm(sqrt(y) ~ as.factor(week)*as.factor(water), data=barley)
> anova(lmadd1,lmfull2)
```

Analysis of Variance Table

Model 1: sqrt(y) ~ week + water

Model 2: sqrt(y) ~ as.factor(week) * as.factor(water)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	27	22.225				
2	20	18.428	7	3.7969	0.5887	0.7574

The large P -value shows that the linear model fits the data as good as the more general effects model, so the linear model is adequate. There is no need to consider higher order polynomials.

How to Include Interaction in a Polynomial Model?

Observe the polynomial model is additive.

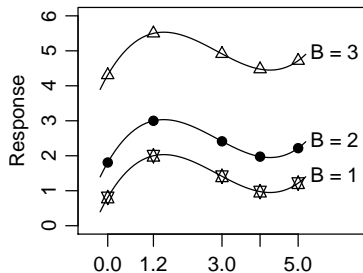
$$y_{ijk} = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_m x_i^m + \beta_j + \varepsilon_{ijk}$$

Factor B only affect the intercept $\theta_0 + \beta_j$ of the polynomial, but not other coefficients.

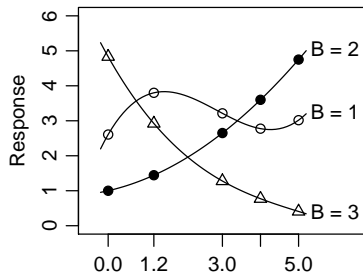
To allow interaction, one should allow the coefficients of the polynomial vary with levels of B. The model would be like

$$y_{ijk} = \theta_{0j} + \theta_{1j} x_i + \theta_{2j} x_i^2 + \cdots + \theta_{mj} x_i^m + \varepsilon_{ijk}.$$

Additive Model



Interaction Model



```
> lmadd1 = lm(sqrt(y) ~ week + water, data=barley)
```

```
> anova(lmadd1)
```

Analysis of Variance Table

Response: sqrt(y)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
week	1	20.346	20.3464	24.718	3.286e-05 ***
water	1	21.893	21.8930	26.597	1.997e-05 ***
Residuals	27	22.225	0.8231		

```
> lmfull2 = lm(sqrt(y) ~ week*as.factor(water), data=barley)
```

```
> anova(lmadd1,lmfull2)
```

Analysis of Variance Table

Model 1: sqrt(y) ~ week + water

Model 2: sqrt(y) ~ as.factor(week) * as.factor(water)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	27	22.225				
2	20	18.428	7	3.7969	0.5887	0.7574