

## Section 9.2.3 Quantitative Factors

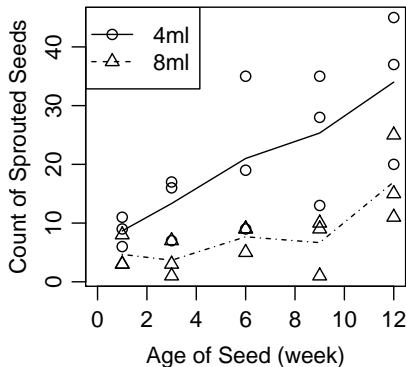
Yibi Huang

## 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)

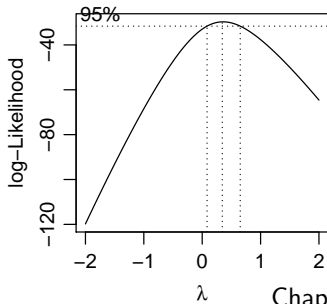
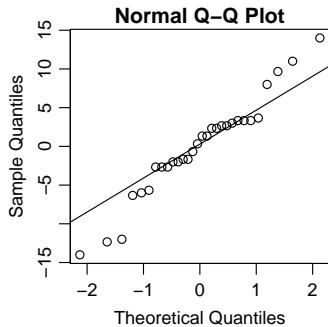
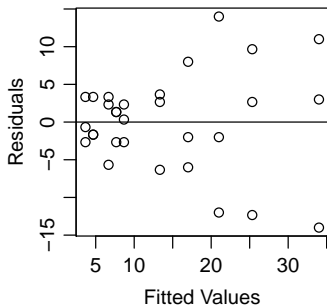
```
> barley = read.table(  
  "http://www.stat.uchicago.edu/~yibi/s222/SproutingBarley.txt",h=T)  
> lmfull = lm(y ~ as.factor(week)*as.factor(water),data=barley)  
> anova(lmfull)
```

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.

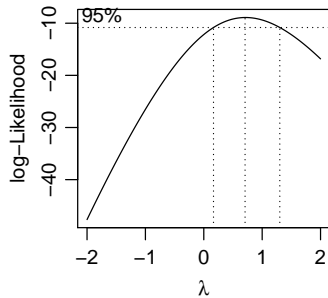
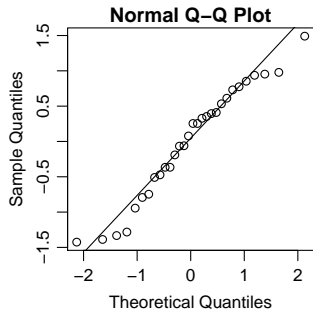
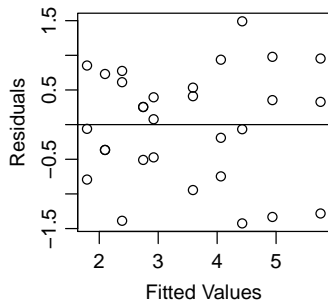
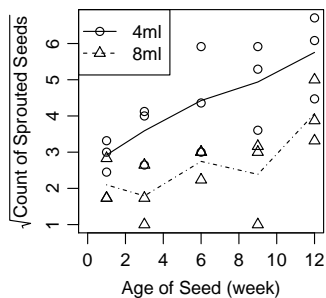
# Sprouting Barley Example Revisit (3)



Spot any problem?

Remedy?

# Sprouting Barley Example Revisit (4)



## Sprouting Barley Example Revisit (4)

```
> lmfull12 = lm(sqrt(y) ~ as.factor(week)*as.factor(water), data=barley)
```

```
> anova(lmfull12)
```

```
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 take quantitative levels of factors into account?

## 9.2.3 Quantitative Factors

Recall in Section 3.10, 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 2-way  $a \times b$  factorial design, of which factor A is quantitative with  $a$  numerical levels  $x_1, \dots, x_a$ , and factor B is categorical, we may consider a polynomial model

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

The polynomial model is nested the additive model

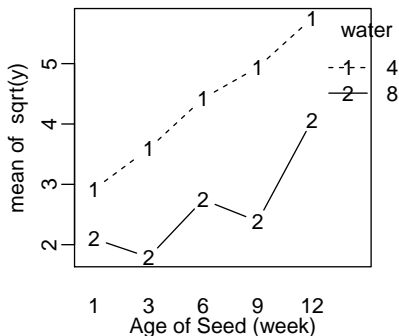
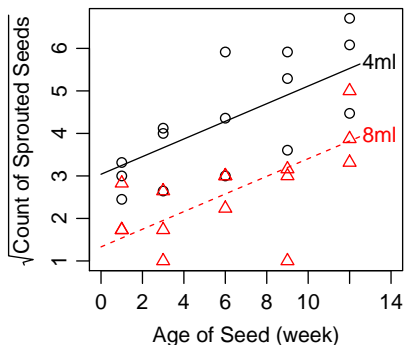
$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk},$$

since one can make

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

- ▶ The order  $m$  of the polynomial **cannot exceed  $a - 1$** . (Why?)

# A Polynomial Model for the Sprouting Barley Data (1)



As **week** is quantitative, the ANOVA table shows no significant interaction, and the square-root transformed response is roughly linearly with **week** from the plot, we thus fit the polynomial model

$$\sqrt{y_{ijk}} = \theta_0 + \theta_1 \text{week}_i + \beta_j + \varepsilon_{ijk}.$$



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:  $\sqrt{y_{ijk}} = \theta_0 + \theta_1 \text{week}_i + \beta_j + \varepsilon_{ijk}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
week	1	20.346	20.3464	24.718	3.286e-05	***
as.factor(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?

## Should We Consider a Higher Order Polynomial Model?

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

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

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

Analysis of Variance Table

```
Model 1: sqrt(y) ~ week + as.factor(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.

## Final Model for the Sprouting Barley Data

From the analysis above, a simplest model with a adequate fit is

$$\sqrt{y_{ijk}} = \theta_0 + \theta_1 \text{week}_i + \beta_j + \varepsilon_{ijk}.$$

```
> lmadd1 = lm(sqrt(y) ~ week + as.factor(water), data=barley)
> lmadd1$coef
      (Intercept)                week as.factor(water)8
      3.0395813          0.2074459          -1.7085268
```

The estimated coefficients are

$$\hat{\theta}_0 \approx 3.0396, \quad \hat{\theta}_1 \approx 0.2074, \quad \hat{\beta}_{8ml} \approx -1.7085$$

and  $\hat{\beta}_{4ml} = 0$  because R use the constraint  $\beta_{4ml} = 0$ .

The fitted model is

$$\begin{aligned} & \sqrt{\text{predicted count of sprouted barley in 100 barley seeds}} \\ & \approx \begin{cases} 3.04 + 0.207(\text{age of seeds in weeks}) & \text{if water} = 4\text{ml} \\ 3.04 - 1.71 + 0.207(\text{age of seeds in weeks}) & \text{if water} = 8\text{ml} \end{cases} \end{aligned}$$

## Final Model for the Sprouting Barley Data

Note that

$y$  = count of sprouted barley in 100 barley seeds  
= percentage of barley seeds sprouted.

$$\begin{aligned}\widehat{\sqrt{y}} &= \sqrt{\text{predicted \% of barley seeds germinated}} \\ &= \begin{cases} 3.04 + 0.207 \text{ week} & \text{if water} = 4\text{ml} \\ 3.04 - 1.71 + 0.207 \text{ week} & \text{if water} = 8\text{ml} \end{cases}\end{aligned}$$

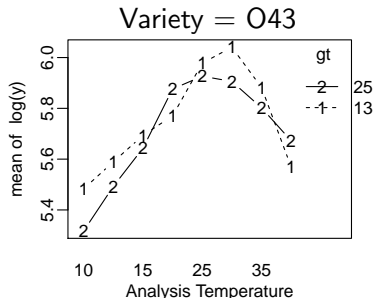
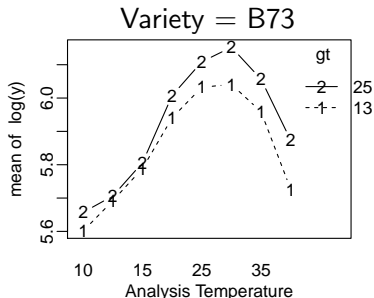
Interpretation: The square-root of the predicted percentage of barley germinated

- ▶ increases by 0.207% for every additional week after harvest, no matter 4 ml or 8 ml of water used in germination
- ▶ decreases by 1.71% if the amount of water is increased from 4 ml to 8 ml, regardless of the age of seeds.

## Example 8.10 Amylase data (Revisit)

$8 \times 2 \times 2$  factorial data with 3 replicates. The factors are

- ▶ Analysis Temperature ( $^{\circ}\text{C}$ ): 10, 13, 15, 20, 25, 30, 35, 40
- ▶ Variety of Maize: B73, O43
- ▶ Growth Temperature:  $13^{\circ}\text{C}$ ,  $25^{\circ}\text{C}$



```
amyl = read.table(  
  "http://www.stat.uchicago.edu/~yibi/s222/amylaze.txt", h=T)  
amyl$at = as.factor(amyl$atemp)  
amyl$gt = as.factor(amyl$gtemp)  
amyl$v = as.factor(amyl$variety)
```

```

> logfit1 = lm(log(y) ~ at*v*gt, data=amyl)
> anova(logfit1)
Response: log(y)

          Df  Sum Sq Mean Sq  F value    Pr(>F)
at          7  3.01613  0.43088   78.8628 < 2.2e-16 ***
v           1  0.58957  0.58957  107.9085 2.305e-15 ***
gt          1  0.00438  0.00438    0.8016 0.3739757
at:v        7  0.02758  0.00394    0.7212 0.6543993
at:gt       7  0.08106  0.01158    2.1195 0.0539203 .
v:gt        1  0.08599  0.08599   15.7392 0.0001863 ***
at:v:gt     7  0.04764  0.00681    1.2457 0.2916176
Residuals  64  0.34967  0.00546

```

The `v:gt` interaction is significant. `at:gt` is at the borderline. Other 2-way and 3-way interactions are not significant. We thus consider the model

$$\log y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \beta\gamma_{jk} + \varepsilon_{ijkl}$$

```
logfit2 = lm(log(y) ~ at + v + gt + v:gt, data=amyl)
```

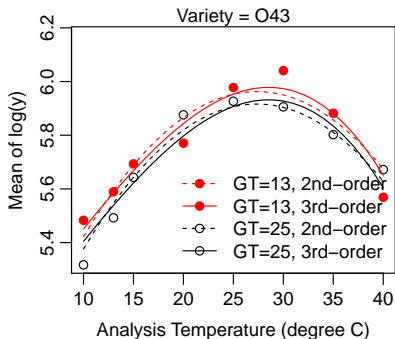
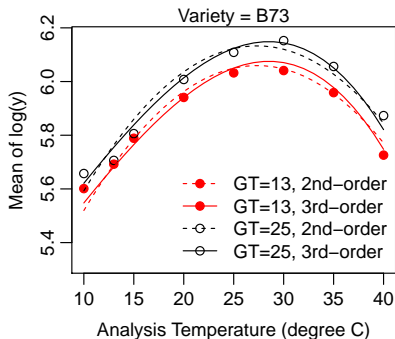
As the analysis temperature  $t$  has 8 numerical levels, we can fit polynomial models up to the 7th-order of the form

$$\log y_{ijkl} = \underbrace{\theta_0 + \theta_1 t_i + \theta_2 t_i^2 + \dots + \theta_m t_i^m}_{=\mu + \alpha_i} + \beta_j + \gamma_k + \beta\gamma_{jk} + \varepsilon_{ijkl}$$

$m$	Model Formula	SSE	$P$ -value	
1	$\theta_0 + \theta_1 t_i$	$+\beta_j + \gamma_k + \beta\gamma_{jk}$	2.64671	$< 10^{-15}$
2	$\theta_0 + \theta_1 t_i + \theta_2 t_i^2$	$+\beta_j + \gamma_k + \beta\gamma_{jk}$	0.55699	0.140
3	$\theta_0 + \theta_1 t_i + \theta_2 t_i^2 + \theta_3 t_i^3$	$+\beta_j + \gamma_k + \beta\gamma_{jk}$	0.51500	0.822
4	$\theta_0 + \theta_1 t_i + \theta_2 t_i^2 + \theta_3 t_i^3 + \theta_4 t_i^4$	$+\beta_j + \gamma_k + \beta\gamma_{jk}$	0.51216	0.791
	$\mu + \alpha_i$	$+\beta_j + \gamma_k + \beta\gamma_{jk}$	0.50595	

Note that all the polynomial models above are nested in the model in the last row of the table. The  $P$ -values in the last column of the table are for comparing these polynomial models with the model in the last row.

We see that the main effects of analysis temperature  $\mu + \alpha_i$  can be described using a 2nd- or a 3rd-order polynomial. No need to consider higher-order polynomials.



- ▶ All 4 curves for the 2nd-order model only differ by the intercept, and are of **the same shape**, ditto for the 3rd-order model, as the models assume no  $at:v$  or  $at:gt$  interactions
- ▶ Observe the GT effect is different for the two varieties — the black curve (GT = 25) is above the red curve (GT = 13) for B73, but not for O43 —  $v:gt$  interactions
- ▶ It's hard to say whether the solid curves (3rd-order) or the dashed curves (2nd-order) fit the data better



## Fitted 2nd-order model (under the baseline constraint)

```
> lm(log(y) ~ atemp+I(atemp^2)+ gt*v, data = amyl)$coef  
(Intercept)      atemp  I(atemp^2)      gt25      vO43  gt25:vO43  
4.7169537  0.0982358 -0.0017969  0.0733670 -0.0968751 -0.1197168
```

$$\widehat{\log y} = 4.717 + 0.09824t - 0.001797t^2 \\ + 0.07337GT25 - 0.0969vO43 - 0.1197GT25 \times vO43$$

where  $y$  = amylase activity,  $t$  = analysis temperature,

$$GT25 = \begin{cases} 0 & \text{if growth temp.} = 13 \\ 1 & \text{if growth temp.} = 25 \end{cases} \quad vO43 = \begin{cases} 0 & \text{for Variety B73} \\ 1 & \text{for Variety O43} \end{cases}$$

The log of amylase activity increases and then decrease with the analysis temperature ( $t$ ) following the polynomial  $C + 0.09824t - 0.001797t^2$ , only the constant  $C$  of the polynomial depends on the variety and the growth temperature (GT).

$$C \approx \begin{cases} 4.717 & \text{for B73 at GT } 13^\circ\text{C} \\ 4.717 + 0.073 = 4.790 & \text{for B73 at GT } 25^\circ\text{C} \\ 4.717 - 0.097 = 4.620 & \text{for O43 at GT } 13^\circ\text{C} \\ 4.717 + 0.073 - 0.097 - 0.1197 \approx 4.573 & \text{for O43 at GT } 25^\circ\text{C} \end{cases}$$

## Interpretation of the 2nd-order model (under the baseline constraint)

$$\widehat{\log y} = C + 0.09824t - 0.001797t^2,$$

where

$$C \approx \begin{cases} 4.717 & \text{for B73 at GT } 13^\circ\text{C} \\ 4.717 + 0.073 = 4.790 & \text{for B73 at GT } 25^\circ\text{C} \\ 4.717 - 0.097 = 4.620 & \text{for O43 at GT } 13^\circ\text{C} \\ 4.717 + 0.073 - 0.097 - 0.120 = 4.573 & \text{for O43 at GT } 25^\circ\text{C} \end{cases}$$

**Variety effect:** At GT  $13^\circ\text{C}$ ,  $25^\circ\text{C}$ ,  $\log(\text{amylase activity})$  for Variety O43 is

$0.097$   
 $0.097 + 0.120 \approx 0.217$  lower than that for Variety B73, i.e., the amylase activity for B73 is  $e^{0.097} \approx 1.10$  times that for O43.  
 $e^{0.217} \approx 1.24$

**Growth Temp effect:** For  $\frac{\text{B73}}{\text{O43}}$  grown at  $25^\circ\text{C}$ , amylase activity is

$e^{0.073} \approx 1.076$   
 $e^{0.073-0.120} \approx e^{-0.047} \approx 0.954$  times as high as those grown at  $\text{GT} = 13^\circ\text{C}$ ,

## Fitted 2nd-order model (under the baseline constraint)

```
> summary(lm(log(y) ~ atemp+I(atemp^2)+ gt*v, data = amyl))
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.717e+00  5.527e-02  85.350 < 2e-16 ***
atemp        9.824e-02  4.899e-03  20.053 < 2e-16 ***
I(atemp^2)  -1.797e-03  9.779e-05 -18.376 < 2e-16 ***
gt25         7.337e-02  2.271e-02   3.231 0.001726 **
v043        -9.688e-02  2.271e-02  -4.266 4.91e-05 ***
gt25:v043   -1.197e-01  3.212e-02  -3.728 0.000337 ***
---
Residual standard error: 0.07867 on 90 degrees of freedom
Multiple R-squared:  0.8674, Adjusted R-squared:  0.8601
F-statistic: 117.8 on 5 and 90 DF,  p-value: < 2.2e-16
```

## Fitted 3rd-order model (under the baseline constraint)

```
> summary(lm(log(y) ~ atemp+I(atemp^2)+I(atemp^3) + gt*v, data = amy1))
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.083e+00  1.459e-01  34.844 < 2e-16 ***
atemp        4.378e-02  2.076e-02   2.109 0.037752 *
I(atemp^2)   5.904e-04  8.912e-04   0.662 0.509406
I(atemp^3)  -3.173e-05  1.178e-05  -2.694 0.008440 **
gt25         7.337e-02  2.196e-02   3.341 0.001221 **
v043        -9.688e-02  2.196e-02  -4.412 2.87e-05 ***
gt25:v043   -1.197e-01  3.106e-02  -3.855 0.000218 ***
---
Residual standard error: 0.07607 on 89 degrees of freedom
Multiple R-squared:  0.8774, Adjusted R-squared:  0.8692
F-statistic: 106.2 on 6 and 89 DF,  p-value: < 2.2e-16
```

- ▶ Only slightly improves over the 2nd-order model
- ▶ The coefficients for `gt25`, `v043`, and `gt25:v043` are identical to those for the 2nd-order model. This is because the data is **balanced**.

## How to Allow Interactions in a Polynomial Model?

Observe the polynomial model below is **additive**. Factor B only affects the intercept  $\theta_0 + \beta_j$  of the polynomial, but not other coefficients.

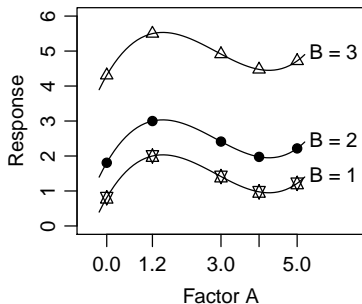
$$y_{ijk} = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_m x_i^m + \beta_j + \varepsilon_{ijk}$$

To allow interaction, one can allow the coefficients of the polynomial vary with levels of B,

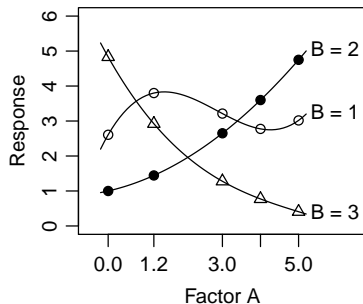
$$y_{ijk} = \theta_{0j} + \theta_{1j} x_i + \theta_{2j} x_i^2 + \cdots + \theta_{mj} x_i^m + \varepsilon_{ijk},$$

i.e., the polynomials for different levels of B can be of different shapes.

Additive Model



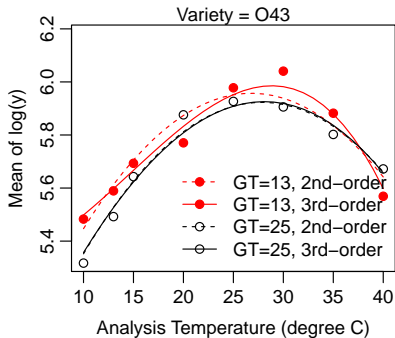
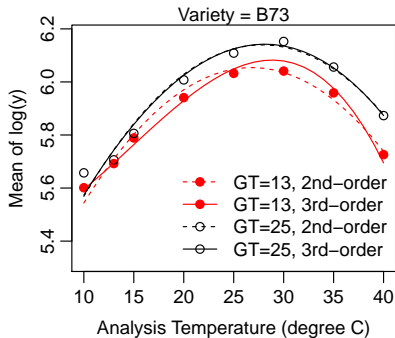
Interaction Model



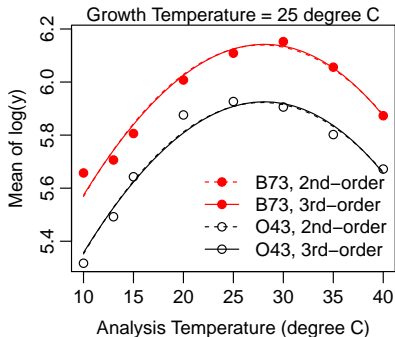
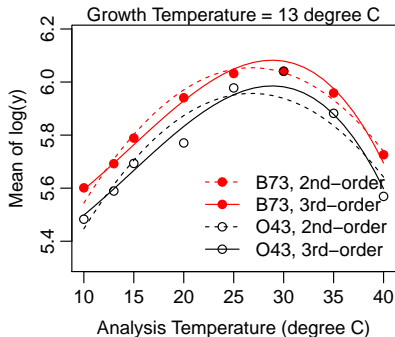
For the amylase data, recall the significance of the two-way interaction `at:gt` is at the borderline with a  $P$ -value of 0.054. We thus consider polynomial models that allows the coefficients of the polynomial to be different at the two growth temperatures

$$\log y_{ijkl} = \underbrace{\theta_{0k} + \theta_{1k}t_i + \theta_{2k}t_i^2 + \dots + \theta_{mk}t_i^m}_{=\mu + \alpha_i + \gamma_k + \alpha\gamma_{ik}} + \beta_j + \beta\gamma_{jk} + \varepsilon_{ijkl}$$

$m$	Model Formula	SSE	$P$ -value	
1	$\theta_0 + \theta_{1k}t_i$	$+\beta_j + \beta\gamma_{jk}$	2.61128	$< 10^{-15}$
2	$\theta_0 + \theta_{1k}t_i + \theta_{2k}t_i^2$	$+\beta_j + \beta\gamma_{jk}$	0.52148	0.0796
3	$\theta_0 + \theta_{1k}t_i + \theta_{2k}t_i^2 + \theta_{3k}t_i^3$	$+\beta_j + \beta\gamma_{jk}$	0.45037	0.7876
4	$\theta_0 + \theta_{1k}t_i + \theta_{2k}t_i^2 + \theta_{3k}t_i^3 + \theta_{4k}t_i^4$	$+\beta_j + \beta\gamma_{jk}$	0.44132	0.788
	$\mu + \alpha_i + \gamma_k + \alpha\gamma_{ik}$	$+\beta_j + \beta\gamma_{jk}$	0.42489	



- ▶ With `at:gt` interactions, the polynomials for  $GT = 13$  and  $GT = 25$  are of different shapes.
- ▶ `v:gt` interactions are present as fitted curves at  $GT = 25$  are above those at  $GT = 13$  for B73, but the other way around for O43.



- ▶ At the same GT, the polynomials for B73 and O43 are identical in shape as this model doesn't include `at:v` interactions.
- ▶ The fits for the 2nd- and the 3rd-order models are nearly identical for  $GT = 25$ , but the two fits are quite different for  $GT = 13$ . The 3rd-order fits are closer to the data points.
- ▶ A 2nd-order polynomial is always symmetric about its peak, but a 3rd-order polynomial may not be symmetric



# Fitted 3rd-Order Model w/ at:gt Interactions

under the baseline constraint

```
> lm(log(y)~(atemp+I(atemp^2)+I(atemp^3))*gt+gt*v, data = aml)$coef
(Intercept)      atemp      I(atemp^2)      I(atemp^3)      gt25
  5.4255022  -0.0028636    0.0025663    -0.0000582   -0.6124712
      v043  atemp:gt25 I(atemp^2):gt25 I(atemp^3):gt25  gt25:v043
 -0.0968751  0.0932958   -0.0039518    0.0000528   -0.1197168
```

At  $GT = 13$ , all terms involving `gt25` are 0. The polynomial is

$$C - 0.002864t_i + 0.002566t_i^2 - 0.0000582t_i^3,$$

where

$$C \approx \begin{cases} 5.4255 & \text{for B73} \\ 5.4255 - 0.0969 \approx 5.3286 & \text{for O43} \end{cases}$$

## Fitted 3rd-Order Model w/ at:gt Interactions

under the baseline constraint

```
> lm(log(y)~(atemp+I(atemp^2)+I(atemp^3))*gt+gt*v, data = aml)$coef
(Intercept)      atemp      I(atemp^2)      I(atemp^3)      gt25
  5.4255022  -0.0028636    0.0025663   -0.0000582  -0.6124712
      v043  atemp:gt25 I(atemp^2):gt25 I(atemp^3):gt25  gt25:v043
 -0.0968751   0.0932958   -0.0039518    0.0000528  -0.1197168
```

At GT = 25, the polynomial is

$$\begin{aligned} & C + (-0.002864 + 0.093296)t + (0.002566 - 0.003952)t^2 \\ & \quad + (-0.0000582 + 0.0000528)t^3 \\ \approx & C + 0.090432t - 0.001386t^2 - 0.0000054t^3 \end{aligned}$$

where

$$C \approx \begin{cases} 5.4255 - 0.6125 = 4.8130 & \text{for B73} \\ 5.4255 - 0.6125 - 0.0969 - 0.1197 \approx 4.5964 & \text{for O43} \end{cases}$$

# Fitted 3rd-Order Model w/ at:gt Interactions

under the baseline constraint

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	5.426e+00	1.957e-01	27.725	< 2e-16	***
atemp	-2.864e-03	2.793e-02	-0.103	0.918577	
I(atemp^2)	2.566e-03	1.199e-03	2.140	0.035156	*
I(atemp^3)	-5.816e-05	1.585e-05	-3.670	0.000421	***
gt25	-6.125e-01	2.767e-01	-2.213	0.029537	*
v043	-9.688e-02	2.089e-02	-4.637	1.25e-05	***
atemp:gt25	9.330e-02	3.950e-02	2.362	0.020438	*
I(atemp^2):gt25	-3.952e-03	1.696e-03	-2.331	0.022115	*
I(atemp^3):gt25	5.285e-05	2.241e-05	2.358	0.020655	*
gt25:v043	-1.197e-01	2.954e-02	-4.052	0.000111	***

---

Residual standard error: 0.07237 on 86 degrees of freedom

Multiple R-squared: 0.8928, Adjusted R-squared: 0.8816

F-statistic: 79.6 on 9 and 86 DF, p-value: < 2.2e-16

## Goodness of Fit

2nd-order model, no `at:gt` interactions

```
> lm1 = lm(log(y) ~ at*v*gt, data=amy1)
> lm22 = lm(log(y) ~ atemp + I(atemp^2)+ v*gt, data=amy1)
> anova(lm22,lm1)
```

Model 1:  $\log(y) \sim \text{atemp} + I(\text{atemp}^2) + \text{gt} * v$

Model 2:  $\log(y) \sim \text{at} * v * \text{gt}$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	90	0.55699				
2	64	0.34967	26	0.20732	1.4595	0.1118

3rd-order model, with `at:gt` interactions

```
> lm33 = lm(log(y)~(atemp + I(atemp^2)+I(atemp^3))*gt+ v*gt, data=amy1)
> anova(lm33,lm1)
```

Model 1:  $\log(y) \sim (\text{atemp} + I(\text{atemp}^2) + I(\text{atemp}^3)) * \text{gt} + v * \text{gt}$

Model 2:  $\log(y) \sim \text{at} * v * \text{gt}$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	86	0.45037				
2	64	0.34967	22	0.1007	0.8378	0.6695

Both models are acceptable.

```
> anova(lm(log(y) ~ at + v*gt, data=amyl))
```

```
Response: log(y)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
at	7	3.01613	0.43088	72.3869	< 2.2e-16 ***
v	1	0.58957	0.58957	99.0475	6.413e-16 ***
gt	1	0.00438	0.00438	0.7358	0.3934347
v:gt	1	0.08599	0.08599	14.4468	0.0002704 ***
Residuals	85	0.50595	0.00595		

```
> anova(lm(log(y) ~ atemp + I(atemp^2)+ v*gt, data=amyl))
```

```
Response: log(y)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
atemp	1	0.87537	0.87537	141.4437	< 2.2e-16 ***
I(atemp^2)	1	2.08972	2.08972	337.6596	< 2.2e-16 ***
v	1	0.58957	0.58957	95.2636	8.996e-16 ***
gt	1	0.00438	0.00438	0.7077	0.4024542
v:gt	1	0.08599	0.08599	13.8948	0.0003374 ***
Residuals	90	0.55699	0.00619		

- ▶ Observe the SS for **v**, **gt**, and **v:gt** in the two ANOVA tables are identical. This is because the data are **balanced**
- ▶ The first model has 5 more parameters than the second model

```

> anova(lm(log(y) ~ at*gt + v*gt, data=amyl))
Response: log(y)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
at	7	3.01613	0.43088	79.0981	< 2.2e-16 ***
gt	1	0.00438	0.00438	0.8040	0.3726670
v	1	0.58957	0.58957	108.2305	< 2.2e-16 ***
at:gt	7	0.08106	0.01158	2.1258	0.0503809 .
gt:v	1	0.08599	0.08599	15.7861	0.0001571 ***
Residuals	78	0.42489	0.00545		

```

> anova(lm(log(y)~(atemp + I(atemp^2)+I(atemp^3))*gt+ v*gt, data=amyl))
Response: log(y)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
atemp	1	0.87537	0.87537	167.1553	< 2.2e-16 ***
I(atemp^2)	1	2.08972	2.08972	399.0393	< 2.2e-16 ***
I(atemp^3)	1	0.04199	0.04199	8.0188	0.0057643 **
gt	1	0.00438	0.00438	0.8363	0.3630174
v	1	0.58957	0.58957	112.5806	< 2.2e-16 ***
atemp:gt	1	0.03543	0.03543	6.7654	0.0109406 *
I(atemp^2):gt	1	0.00009	0.00009	0.0170	0.8965610
I(atemp^3):gt	1	0.02911	0.02911	5.5590	0.0206550 *
gt:v	1	0.08599	0.08599	16.4206	0.0001109 ***
Residuals	86	0.45037	0.00524		

- ▶ The SS for **gt**, **v**, and **gt:v** in the two ANOVA tables are identical since the data are **balanced**
- ▶ The first model has 8 more parameters than the second model