The 2^3 Factorial Design

The 2^3 Factorial Design

Evaluate the effects:

- **Temperature** (Factor A);
- **Substrate concentration** (Factor B) and
- **Type of enzyme** (Factor C)

on the yield of an enzymatic reaction.

Photo by Polina Tankilevitch from Pexels

$k = 3$ factors

Factor	Low level	High level
Factor A: Temperature	20 ˚C	30 ˚C
Factor B: Substrate	20 g/L	40 g/L
Factor C: Enzyme	E1	E2

numerical

$\Rightarrow$ categorical

The 2^3 Factorial Design: Building the Experimental Matrix

Experimental Matrix:

	Factor A x_A	Factor B x_B	Factor C x_C	Treatment
1				
2				
3				
4				
5				
6				
7				
8				

1) Determine the number of treatments:

$$2^k = 2^3 = 8 \text{ treatments}$$

number of factors = 3

The 2^3 Factorial Design: Building the Experimental Matrix

Experimental Matrix:

	Factor A x_A	Factor B x_B	Factor C x_C	Treatment
1	-1	-1		(1)
2	+1	-1		a
3	-1	+1		b
4	+1	+1		ab
5				
6				
7				
8				

1) Determine the number of treatments:

$$2^k = 2^3 = 8 \text{ treatments}$$

2) Build a 2^2 design in the standard order

$$(1) \quad a \quad b \quad ab$$

The 2^3 Factorial Design: Building the Experimental Matrix

Experimental Matrix:

	Factor A x_A	Factor B x_B	Factor C x_C	Treatment
1	-1	-1	-1	(1)
2	+1	-1	-1	a
3	-1	+1	-1	b
4	+1	+1	-1	ab
5	-1	-1	+1	
6	+1	-1	+1	
7	-1	+1	+1	
8	+1	+1	+1	

1) Determine the number of treatments:

$$2^k = 2^3 = 8 \text{ treatments}$$

2) Build a 2^2 design in the standard order

$$(1) \quad a \quad b \quad ab$$

3) Double the 2^2 design assigning:

$x_C = -1$ for the upper half and
$x_C = +1$ for the lower half

The 2^3 Factorial Design: Building the Experimental Matrix

Experimental Matrix:

	Factor A x_A	Factor B x_B	Factor C x_C	Treatn
1	-1	-1	-1	(1)
2	+1	-1	-1	a
3	-1	+1	-1	b
4	+1	+1	-1	ab
5	-1	-1	+1	c
6	+1	-1	+1	ac
7	-1	+1	+1	bc
8	+1	+1	+1	abc

Treatments are denominated by the factors tested at their high levels

umber of treatments:

8 treatments

2) Build a 2^2 design in the standard order

$$(1) \quad a \quad b \quad ab$$

3) Double the 2^2 design assigning:

$x_C = -1$ for the upper half and
$x_C = +1$ for the lower half

4) Assign the treatment names

The 2^3 Factorial Design: Building the Experimental Matrix

Experimental Matrix:

	Factor A x_A	Factor B x_B	Factor C x_C	Treatment
1	-1	-1	-1	(1)
2	+1	-1	-1	a
3	-1	+1	-1	b
4	+1	+1	-1	ab
5	-1	-1	+1	c
6	+1	-1	+1	ac
7	-1	+1	+1	bc
8	+1	+1	+1	abc

the number of treatments:

This table presents the standard order of the experimental matrix for a 2^3 factorial design.

$(1) \quad a \quad b \quad ab$

3) Double the 2^2 design assigning:

$x_C = -1$ for the upper half and
$x_C = +1$ for the lower half

4) Assign the treatment names

The 2^3 Factorial Design: Assigning the Natural Variables

Experimental Matrix:

Run #	Factor A x_A	Factor B x_B	Factor C x_C	Treatment	Temperature T (˚C)	Substrate S (g/L)	Enzyme E
1	-1	-1	-1	(1)			
2	+1	-1	-1	a			
3	-1	+1	-1	b			
4	+1	+1	-1	ab			
5	-1	-1	+1	c			
6	+1	-1	+1	ac			
7	-1	+1	+1	bc			
8	+1	+1	+1	abc			

Factor	Low level	High level
A: Temperature	20 ˚C	30 ˚C
B: Substrate	20 g/L	40 g/L
C: Enzyme	E1	E2

The 2^3 Factorial Design: Running the Experiments

Experimental Matrix:

Run #	Factor A x_A	Factor B x_B	Factor C x_C	Treatment	Temperature T (˚C)	Substrate S (g/L)	Enzyme E	Yield (%) $n = 2$ replicates
1	-1	-1	-1	(1)	20	20	E1	
2	+1	-1	-1	a	30	20	E1	
3	-1	+1	-1	b	20	40	E1	
4	+1	+1	-1	ab	30	40	E1	
5	-1	-1	+1	c	20	20	E2	
6	+1	-1	+1	ac	30	20	E2	
7	-1	+1	+1	bc	20	40	E2	
8	+1	+1	+1	abc	30	40	E2	

The 2^3 Factorial Design: Running the Experiments

Experimental Matrix:

Run #	Factor A x_A	Factor B x_B	Factor C x_C	Treatment	Temperature T (˚C)	Substrate S (g/L)	Enzyme E	Yield (%) $n = 2$ replicates	
1	-1	-1	-1	(1)	20	20	E1	59	61
2					30	20	E1	74	70
3					20	40	E1	50	58
4					30	40	E1	69	67
5	-1	-1	+1	c	20	20	E2	50	54
6	+1	-1	+1	ac				81	85
7	-1	+1	+1	bc				46	44
8	+1	+1	+1	abc	30	40	E2	79	81

The 16 runs were performed in a random order:
completely randomized experiment

$$N = 2^k \times n = 2^3 \times 2 = 16$$

The 2^3 Factorial Design: Analysis of Variance

Source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
Factor A	SS_A	1	$MS_A = SS_A$	$F_0 = \dfrac{MS_i}{MS_E}$	
Factor B main factors	SS_B	1	$MS_B = SS_B$		
Factor C	SS_C	1	$MS_C = SS_C$		
AB	SS_{AB}	1	$MS_{AB} = SS_{AB}$		
AC two-factor interactions	SS_{AC}	1	$MS_{AC} = SS_{AC}$		
BC	SS_{BC}	1	$MS_{BC} = SS_{BC}$		
ABC three-factor interaction	SS_{ABC}	1	$MS_{ABC} = SS_{ABC}$		
Error	SS_E	$2^k(n-1)$	$MS_E = \dfrac{SS_E}{2^k(n-1)}$		
Total	SS_T	$2^k \times n - 1$			

The 2^3 Factorial Design: Analysis of Variance

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value
Temperature	2116	1	2116	264.5	0.0000
Substrate	100	1	100	12.5	0.0077
Enzyme	9	1	9	1.125	0.3198
Temp x Subst	9	1	9	1.125	0.3198
Temp x Enz	400	1	400	50	0.0001
Subst x Enz	0	1	0	0	1.0000
Temp x Subs x Enz	1	1	1	0.125	0.7328
Error	64	8	8		
Total	2699	15			

$F_{1,8} = 5.318$

The 2^3 Factorial Design: Analysis of Variance

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p-value	
Temperature	2116	1	2116	264.5	0.0000	significant
Substrate	100	1	100	12.5	0.0077	significant
Enzyme	9	1	9	1.125	0.3198	
Temp x Subst	9	1	9	1.125	0.3198	
Temp x Enz	400	1	400	50	0.0001	significant
Subst x Enz	0	1	0	0	1.0000	
Temp x Subs x Enz	1	1	1	0.125	0.7328	
Error	64	8	8			
Total	2699	15				

$F_{1,8} = 5.318$

The 2^3 Factorial Design: Regression Model

$$y_{pred} = \underbrace{b_0 + b_A x_A + b_B x_B + b_C x_C}_{1^{st}\text{ order linear}} + \underbrace{b_{AB} x_A x_B + b_{AC} x_A x_C + b_{BC} x_B x_C}_{\text{two-factor interaction}} + \underbrace{b_{ABC} x_A x_B x_C}_{\text{three-factor interaction}}$$

predicted (calculated) response

The 2^3 Factorial Design: Regression Model

$$y_{pred} = b_0 + b_A x_A + b_B x_B + b_C x_C + \cancel{b_{AB} x_A x_B} + b_{AC} x_A x_C + \cancel{b_{BC} x_B x_C} + \cancel{b_{ABC} x_A x_B x_C}$$

Temp x Subs   Subs x Enz   Temp x Subs x Enz

Source	SS	df	MS	F_0	p-value	
Temperature	2116	1	2116	264.5	0.0000	significant
Substrate	100	1	100	12.5	0.0077	significant
Enzyme	9	1	9	1.125	0.3198	
Temp x Subs	9	1	9	1.125	0.3198	
Temp x Enz	400	1	400	50	0.0001	significant
Subs x Enz	0	1	0	0	1.0000	
Temp x Subs x Enz	1	1	1	0.125	0.7328	
Error	64	8	8			
Total	2699	15				

The 2^3 Factorial Design: Regression Model

$$y_{pred} = b_0 + b_A x_A + b_B x_B + b_C x_C + \cancel{b_{AB} x_A x_B} + b_{AC} x_A x_C + \cancel{b_{BC} x_B x_C} + \cancel{b_{ABC} x_A x_B x_C}$$

Enzyme

Temp x Subs

Subs x Enz

Temp x Subs x Enz

Source	SS	df	MS	F_0	p-value
Temperature	2116	1	2116	264.5	0.0000
Substrate	100	1	100	12.5	0.0077
Enzyme	9	1	9	1.125	0.3198
Temp x Subs	9	1	9	1.125	0.3198
Temp x Enz	400	1	400	50	0.0001
Subs x Enz	0	1	0	0	1.0000
Temp x Subs x Enz	1	1	1	0.125	0.7328
Error	64	8	8		
Total	2699	15			

We must keep the enzyme coefficient, since the enzyme is significant.
However, its effect depends on the temperature.

The 2^3 Factorial Design: Regression Model

$$y_{pred} = b_0 + b_A x_A + b_B x_B + b_C x_C + \cancel{b_{AB} x_A x_B} + b_{AC} x_A x_C + \cancel{b_{BC} x_B x_C} + \cancel{b_{ABC} x_A x_B x_C}$$

Enzyme Temp x Subs Subs x Enz Temp x Subs x Enz

Reduced model for coded variables:

$$y_{pred} = b_0 + \underbrace{b_T x_T}_{\text{Temp}} + \underbrace{b_S x_S}_{\text{Subs}} + \underbrace{b_E x_E}_{\text{Enz}} + \underbrace{b_{TE} x_T x_E}_{\text{Temp x Enz}}$$

The 2^3 Factorial Design: Regression Models

Regression Model for Coded Variables:

	Coefficient	p-value
Intercept (b_0)	64.25	$< 2\times10^{-16}$
x_T	11.5	1.9×10^{-9}
x_S	-2.5	0.0027
x_E	0.75	0.27
$x_T x_E$	5.0	9.25×10^{-6}

$R^2 = 0.9726$

Regression Model for Natural Variables:

	Coefficient	p-value
Intercept (b_0)	14.25	0.00341
T	2.3	1.9×10^{-9}
S	-0.25	0.0027
x_E	-24.25	1.5×10^{-5}
$T\times x_E$	1	9.25×10^{-6}

$R^2 = 0.9726$

significant in the ANOVA

The 2^3 Factorial Design: Regression Models

Regression Model for Coded Variables:

	Coefficient	p-value
Intercept (b_0)	64.25	$< 2\times10^{-16}$
x_T	11.5	1.9×10^{-9}
x_S	-2.5	0.0027
x_E	0.75	0.27
$x_T x_E$	5.0	9.25×10^{-6}

$R^2 = 0.9726$

Regression Model for Natural Variables:

	Coefficient	p-value	
Intercept (b_0)	14.25	0.00341	
T	2.3	1.9×10^{-9}	20 30
S	-0.25	0.0027	20 40
x_E	-24.25	1.5×10^{-5}	-1 +1
$T\times x_E$	1	9.25×10^{-6}	

$R^2 = 0.9726$

the order of magnitude
of the natural variables
interferes in the statistical analysis

The 2^3 Factorial Design: Model Adequacy

Final regression model:

$$Y(\%) = 14.25 + 2.3T - 0.25S - 24.25x_E + T \cdot x_E$$

$x_E = -1$ for enzyme E1

$x_E = +1$ for enzyme E2

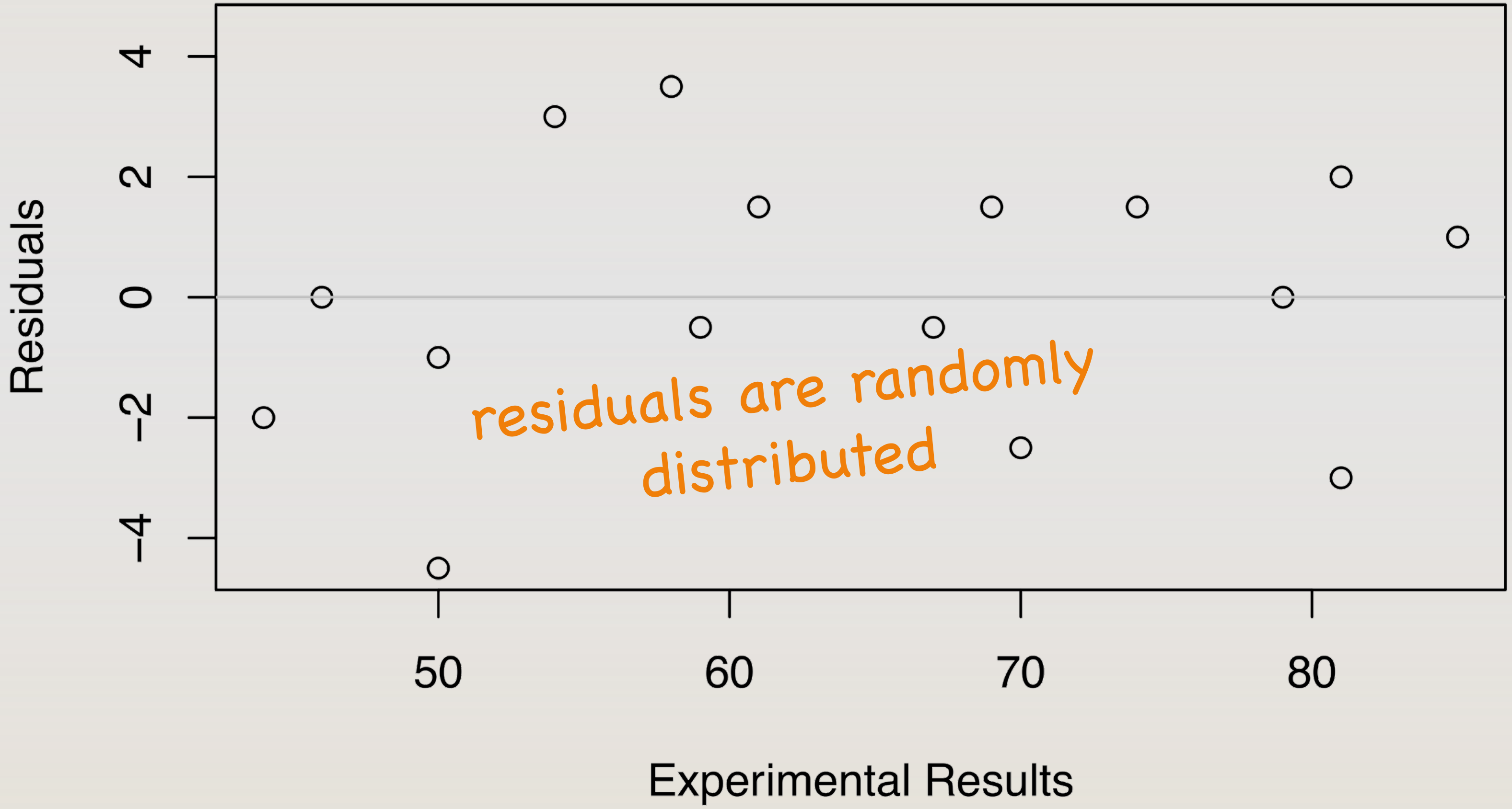
Regression Model for Natural Variables:

	Coefficient	*p*-value
	14.25	0.00341
	2.3	1.9×10^{-9}
	−0.25	0.0027
	−24.25	1.5×10^{-5}
$T\times x_E$	1	9.25×10^{-6}

$R^2 = 0.9726$

97.3 % of the data variability is explained by the model

The 2^3 Factorial Design: Residuals Analysis



The 2^3 Factorial Design: Contour Plot

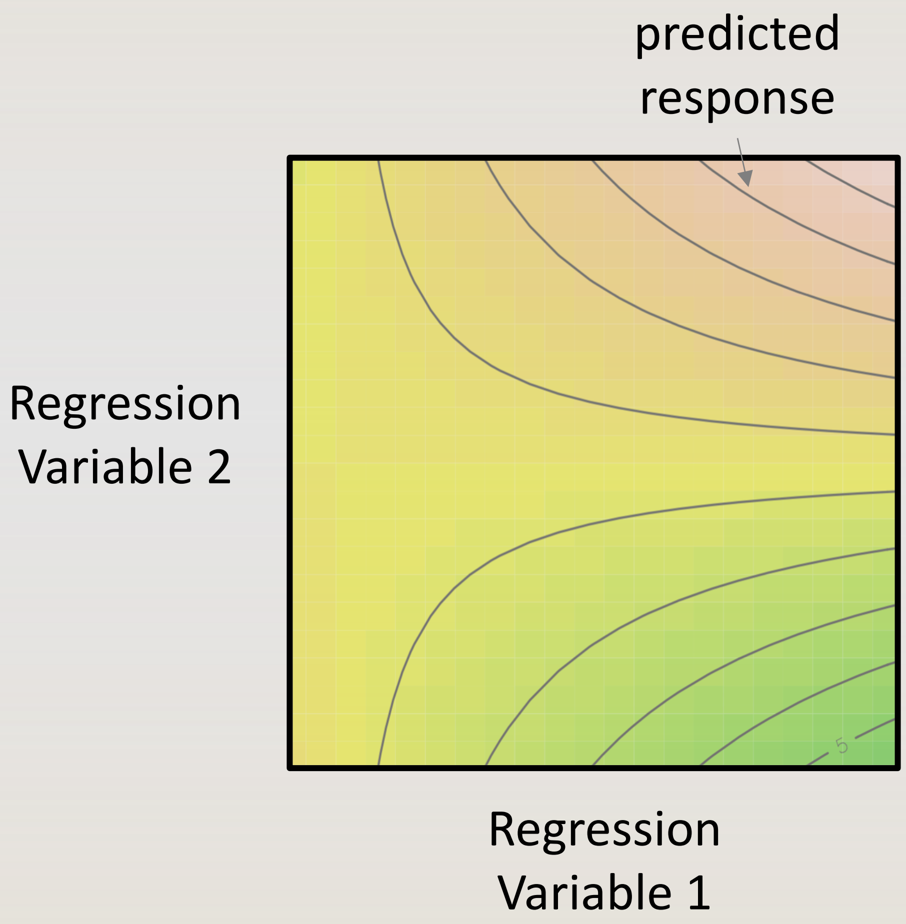
2^3 design: 3 regression variables

Which contour plots are the best to show the results?

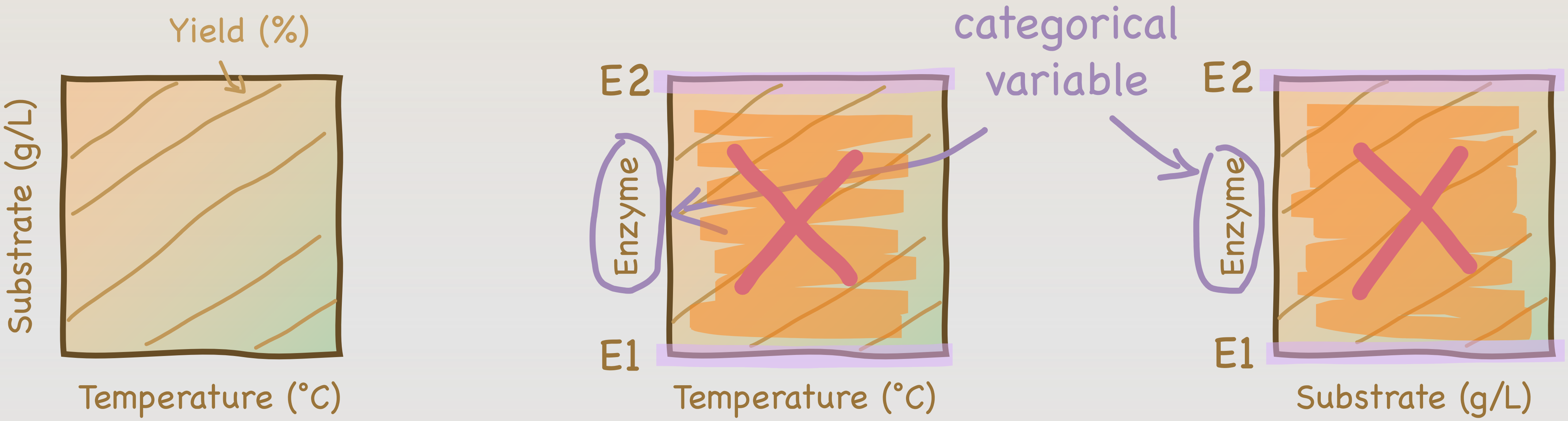
Temperature

Substrate

Enzyme

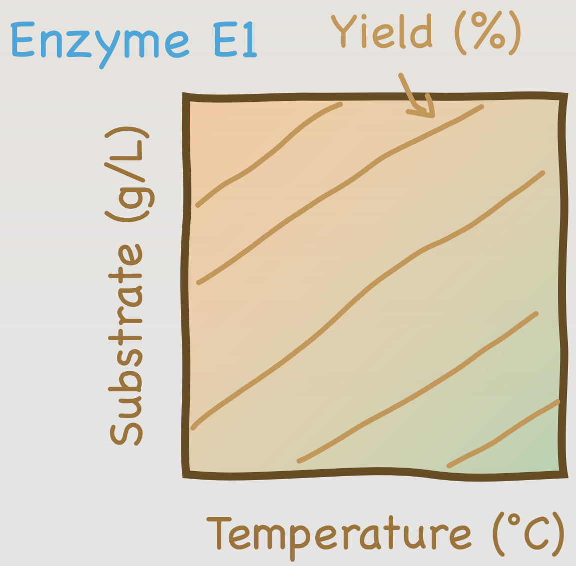


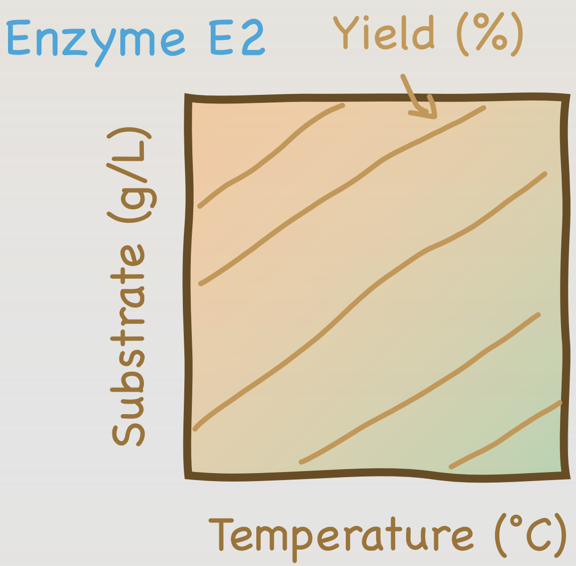
The 2^3 Factorial Design: Contour Plot



A contour plot with a categorical variable in the axis does not make sense.

The 2^3 Factorial Design: Contour Plot

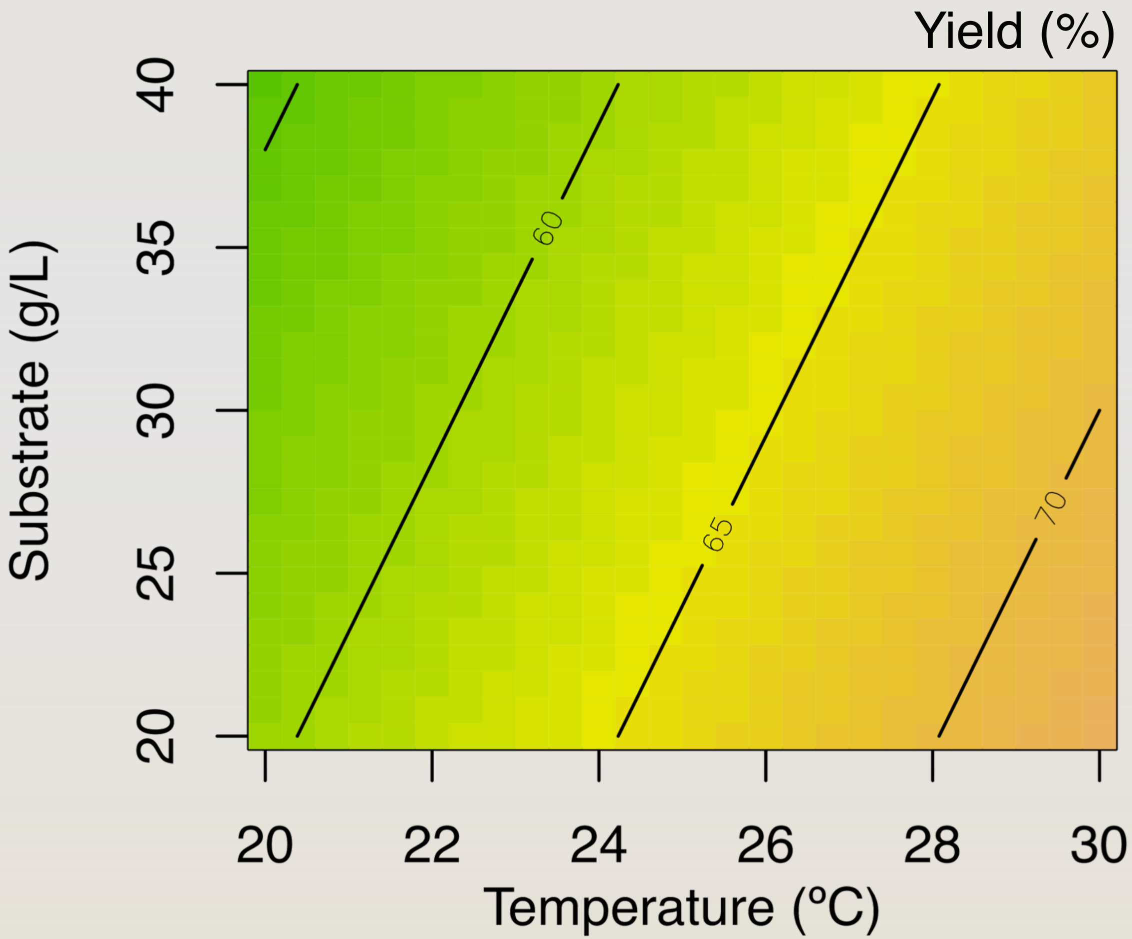




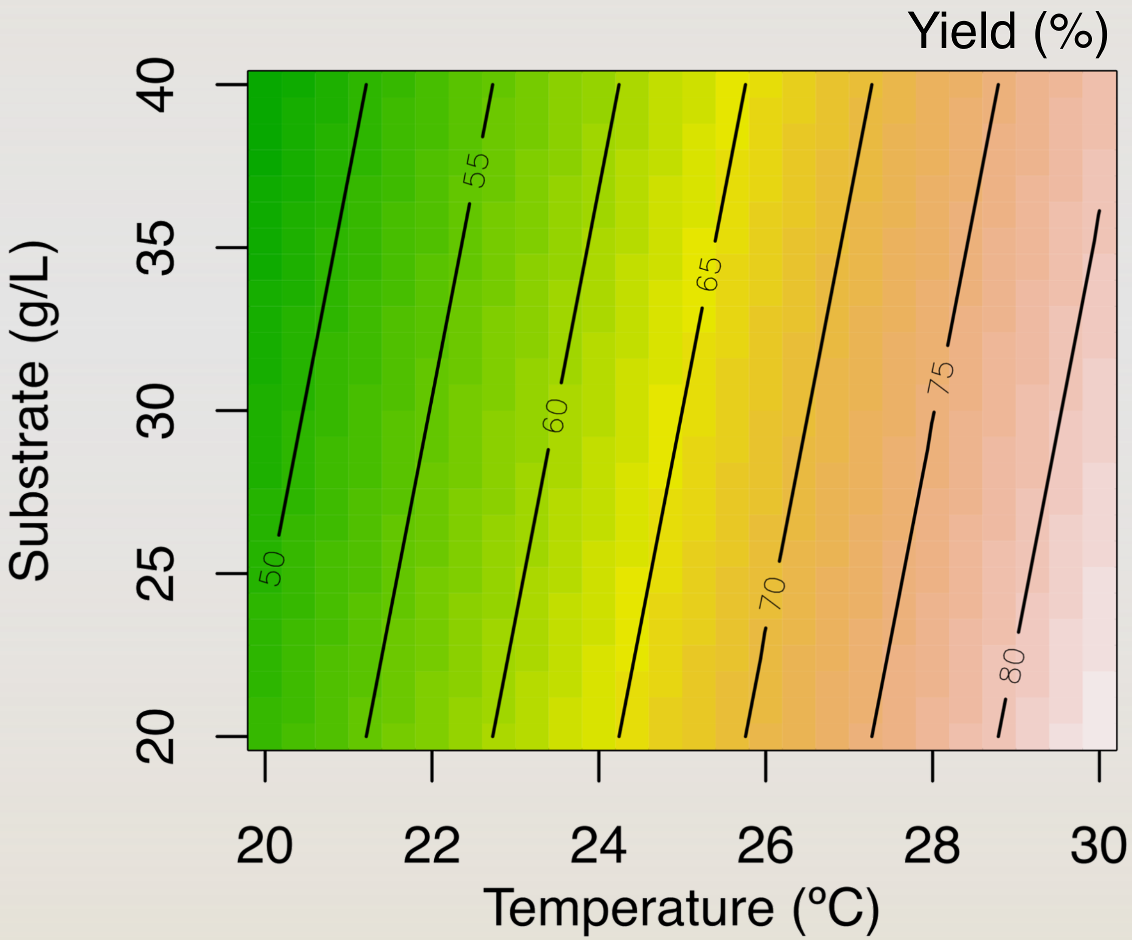
As the enzyme can only assume the values -1 and +1, we are going to build a contour plot Temp x Subs for each enzyme type.

The 2^3 Factorial Design: Contour Plot

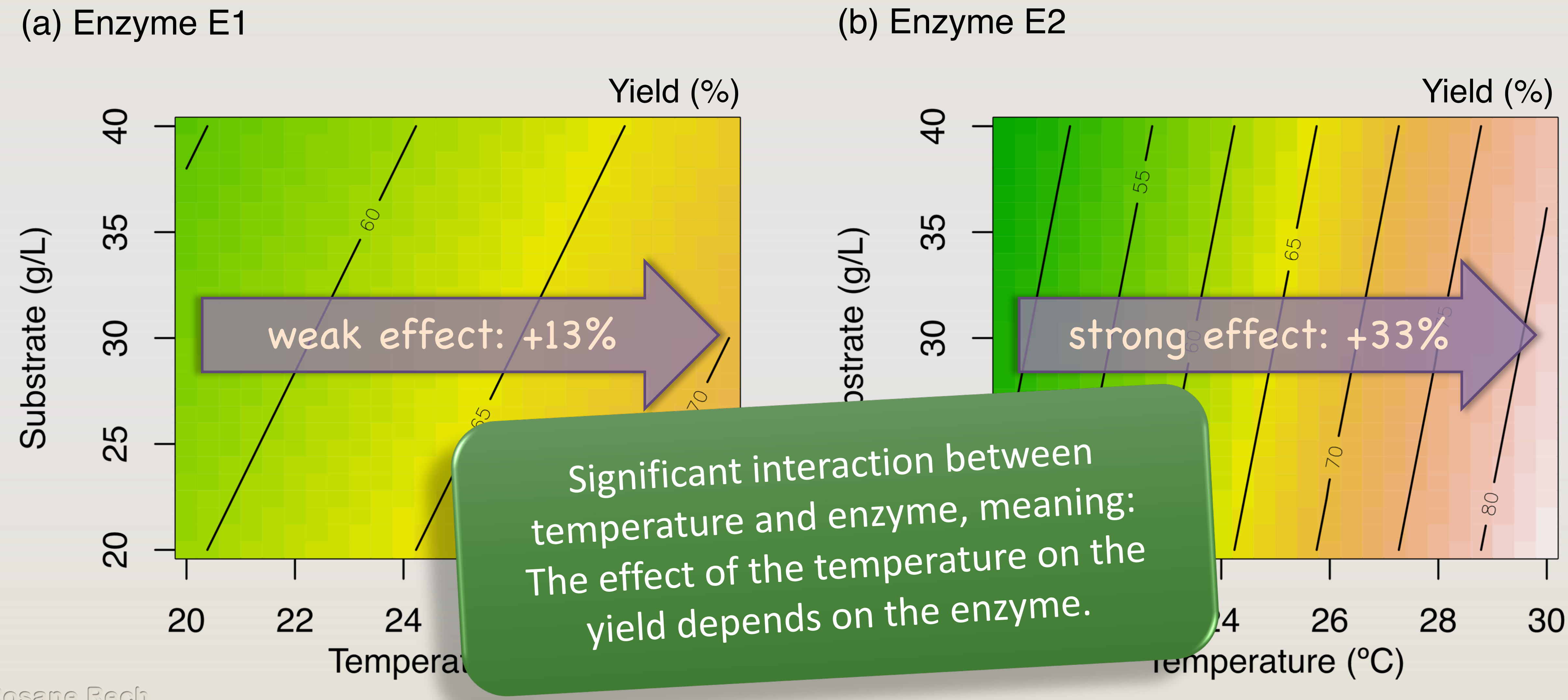
(a) Enzyme E1



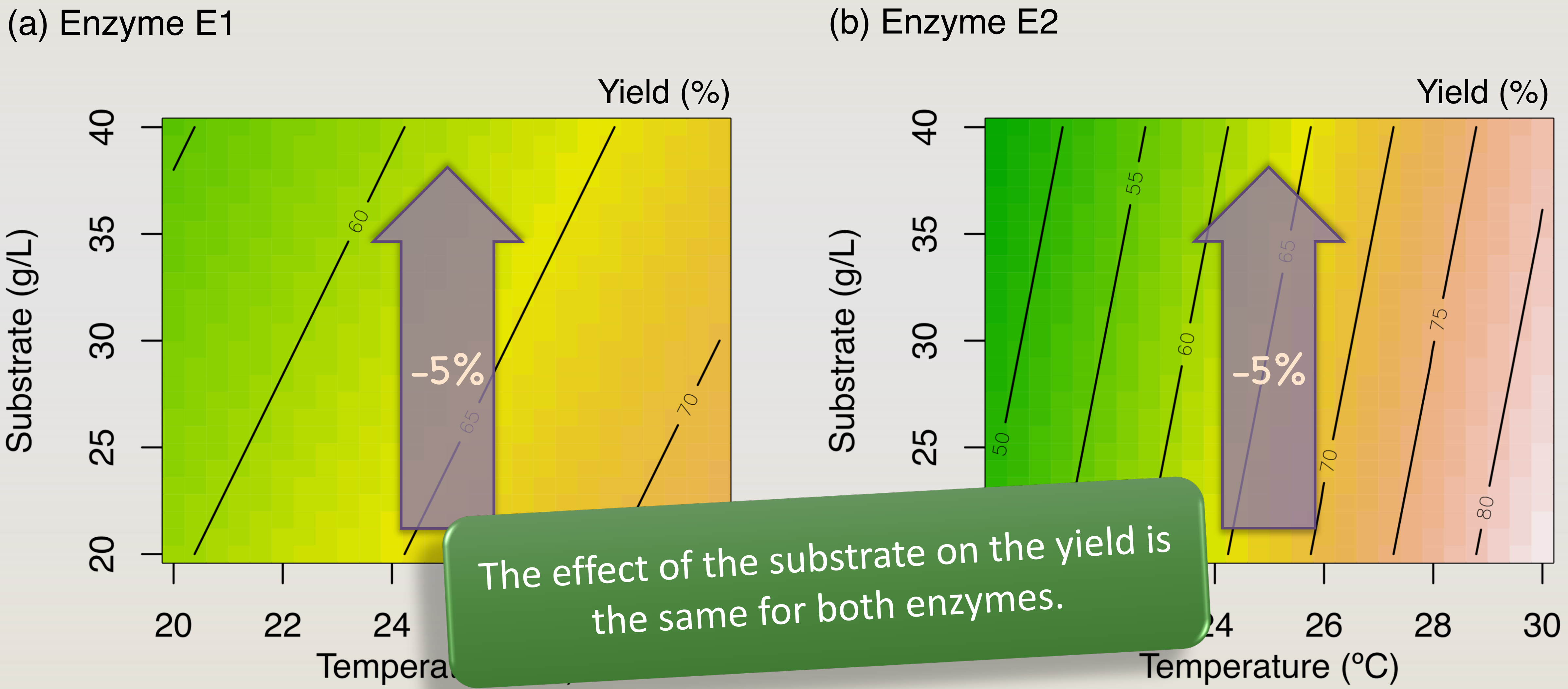
(b) Enzyme E2



The 2^3 Factorial Design: Contour Plot



The 2^3 Factorial Design: Contour Plot



The effect of the substrate on the yield is the same for both enzymes.