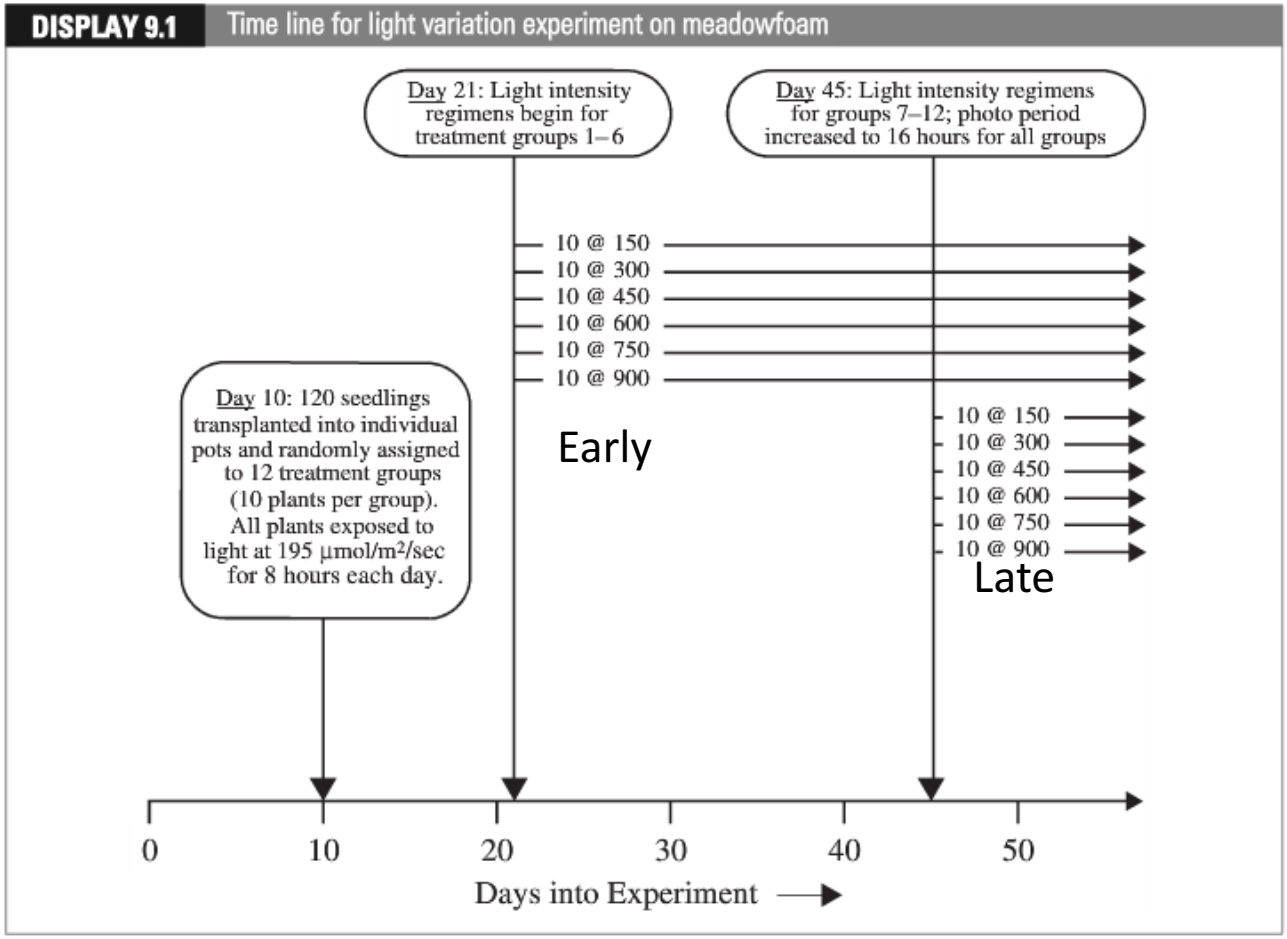
Multiple Regression

OVERVIEW OF A MORE GENERAL REGRESSION MODEL

Meadowfoam Flowering

Meadowfoam Flowering



Meadowfoam Flowering

```
data flower;
input Flowers       Time       Intensity Timing $;
datalines;
62.3        1        150       Late
77.4        1        150       Late
55.3        1        300       Late
54.2        1        300       Late
49.6        1        450       Late
61.9        1        450       Late
39.4        1        600       Late
45.7        1        600       Late
31.3        1        750       Late
44.9        1        750       Late
36.8        1        900       Late
41.9        1        900       Late
77.8        0        150       Early
75.6        0        150       Early
69.1        0        300       Early
78        0        300       Early
57        0        450       Early
71.1        0        450       Early
62.9        0        600       Early
52.2        0        600       Early
60.3        0        750       Early
45.6        0        750       Early
52.6        0        900       Early
44.4        0        900       Early
;
```

If we thought of this question like a one-group problem, there would be 12 treatments:

- Intensity = 150, Timing = Late

....

- Intensity = 900, Timing = Early

(there are two codings of same timing variable)

But, it is more helpful to think of there being all combinations of two grouping variables

- 6 light intensities
- 2 timing levels

Meadowfoam Flowering Flowers vs. Intensity

$\mu\{\text{flowers}|\text{intensity}\} = \beta_0 + \beta_1 \text{intensity}$

```
proc reg data = flower;
model flowers = intensity;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

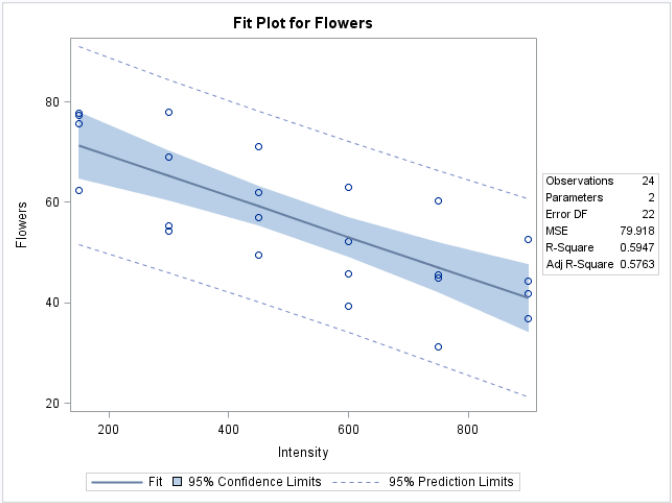
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2579.75004	2579.75004	32.28	<.0001
Error	22	1758.18621	79.91756		
Corrected Total	23	4337.93625			

Root MSE	8.93966	R-Square	0.5947
Dependent Mean	56.13750	Adj R-Sq	0.5763
Coeff Var	15.92458		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	77.38500	4.16119	18.60	<.0001
Intensity	1	-0.04047	0.00712	-5.68	<.0001

	Light Intensity (μmol/m^2/sec)					
	150	300	450	600	750	900
Timing: Late (at PFI)	62.3 77.4	55.3 54.2	49.6 61.9	39.4 45.7	31.3 44.9	36.8 41.9
Timing: Early (24 days before PFI)	77.8 75.6	69.1 78.0	57.0 71.1	62.9 52.2	60.3 45.6	52.6 44.4
Flowers (avg)	73.27	64.15	59.9	50.05	45.53	43.93



Meadowfoam Flowering
Flowers vs. Intensity & Time

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1251.61400	1251.61400	34.44	0.0002
Error	10	363.46267	36.34627		
Corrected Total	11	1615.07667			

Root MSE	6.02879	R-Square	0.7750
Dependent Mean	62.21667	Adj R-Sq	0.7525
Coeff Var	9.68999		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	83.14667	3.96863	20.95	<.0001
Intensity	1	-0.03987	0.00679	-5.87	0.0002

```
proc reg data = flower;
where time = 0;
model flowers = intensity;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1328.71207	1328.71207	26.20	0.0005
Error	10	507.19710	50.71971		
Corrected Total	11	1835.90917			

Root MSE	7.12178	R-Square	0.7237
Dependent Mean	50.05833	Adj R-Sq	0.6961
Coeff Var	14.22696		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	71.62333	4.68813	15.28	<.0001
Intensity	1	-0.04108	0.00803	-5.12	0.0005

```
proc reg data = flower;
where time = 1;
model flowers = intensity;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3466.70045	1733.35023	41.78	<.0001
Error	21	871.23580	41.48742		
Corrected Total	23	4337.93625			

Root MSE	6.44107	R-Square	0.7992
Dependent Mean	56.13750	Adj R-Sq	0.7800
Coeff Var	11.47374		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	83.46417	3.27377	25.49	<.0001
Intensity	1	-0.04047	0.00513	-7.89	<.0001
Time	1	-12.15833	2.62956	-4.62	0.0001

```
proc reg data = flower;
model flowers = intensity time;
run;
```

Meadowfoam Flowering Flowers vs. Intensity & Time

$\mu\{\text{flowers}|\text{intensity,time}\} = \beta_0 + \beta_1 \text{intensity} + \beta_2 \text{time}$

```
proc reg data = flower;
model flowers = intensity time;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

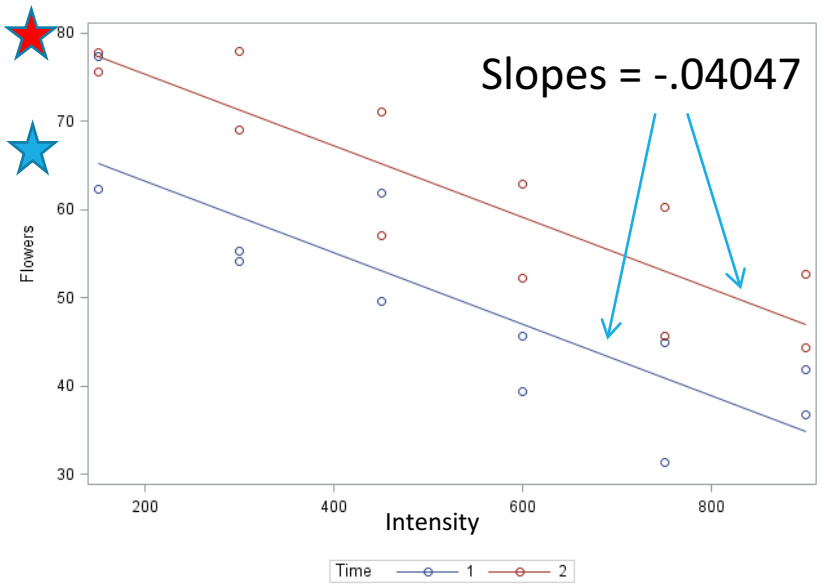
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3466.70045	1733.35023	41.78	<.0001
Error	21	871.23580	41.48742		
Corrected Total	23	4337.93625			

Root MSE	6.44107	R-Square	0.7992
Dependent Mean	56.13750	Adj R-Sq	0.7800
Coeff Var	11.47374		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	83.46417	3.27377	25.49	<.0001
Intensity	1	-0.04047	0.00513	-7.89	<.0001
Time	1	-12.15833	2.62956	-4.62	0.0001

★ Intercept = 83.45417 -12.15833
= 71.3



★ Intercept = 83.46
(Note that X = 0 is off the left end of the chart)

Multiple Regression

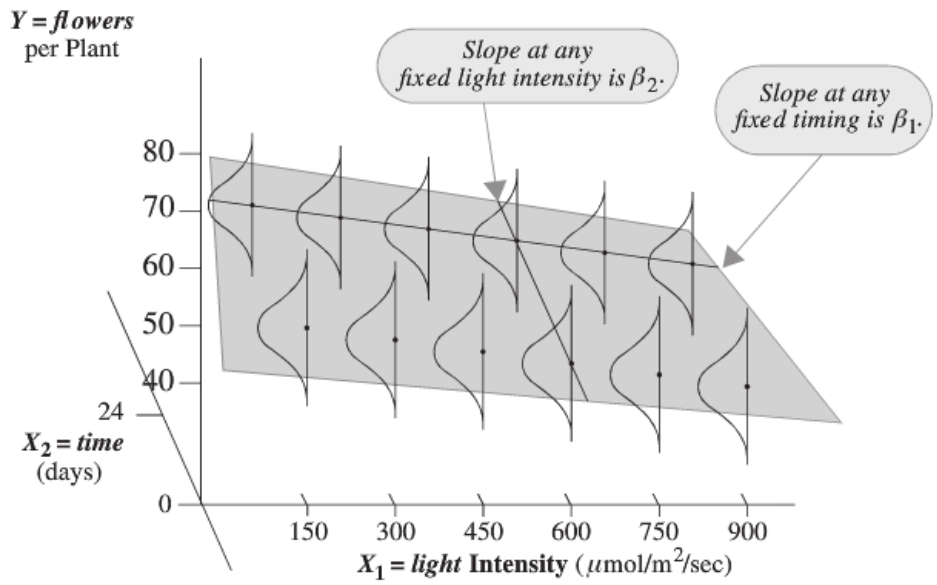
The regression of Y on X_1 and X_2 is $\mu\{Y|X_1, X_2\}$.

Regression plane: $\mu\{flowers|light, time\} = \beta_0 + \beta_1\, light + \beta_2\, time$

$\text{Var}\{flowers \mid light, time\} = \sigma^2$

DISPLAY 9.5 Model for the regression surface of flowers per plant under 12 treatment levels as a regression plane

Regression plane: $\mu\{flowers|light, time\} = \beta_0 + \beta_1\, light + \beta_2\, time$



"Constant Variance Assumption"

This assumption is needed for the inferential tools developed in the next Chapter (intervals, tests, etc.)

Categorical Explanatory Variable

Regression plane: $\mu\{flowers|light, time\} = \beta_0 + \beta_1\, light + \beta_2\, time$

Explanatory Variable Values:

Time: {0: "24 days before PFI", 1: "After PFI"}

Light: {150, 300, 450, 600, 750, 900}

$$\mu\{\text{flowers}|\text{light, time} = 0\} = \beta_0 + \beta_1 \text{light} + \beta_2(0)$$

$$\mu\{\text{flowers}|\text{light, time} = 0\} = (\beta_0) + \beta_1 \text{light}$$

$$\mu\{\text{flowers}|\text{light, time} = 1\} = \beta_0 + \beta_1 \text{light} + \beta_2(1)$$

$$\mu\{\text{flowers}|\text{light, time} = 1\} = (\beta_0 + \beta_2) + \beta_1 \text{light}$$

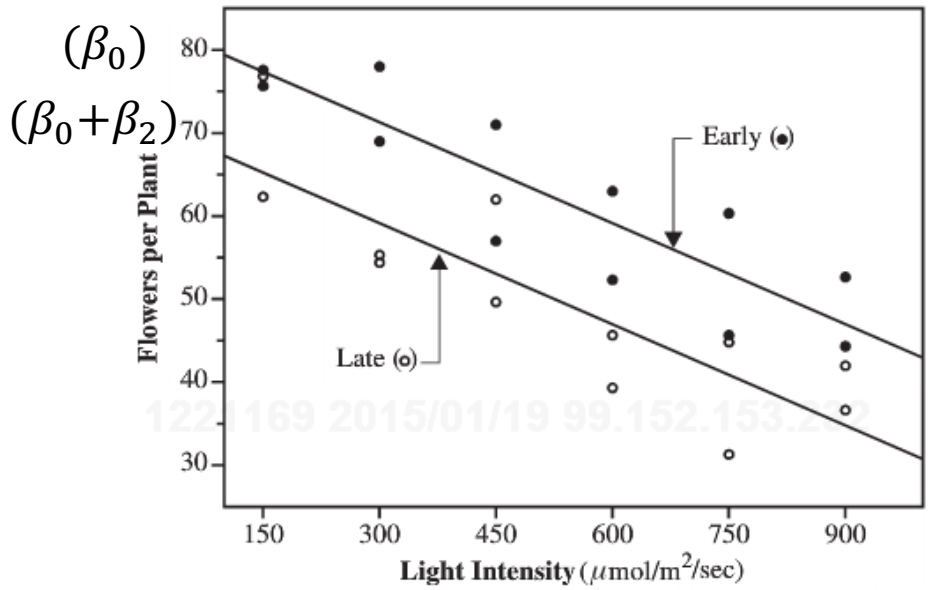
β_2 is the adjustment in the expected (mean) number of flowers in the "after PFI" group relative to the "24 days before PFI' group

Meadowfoam: Categorical Explanatory Variable

$\mu\{\text{flowers}|\text{light, time} = 0\} = \beta_0 + \beta_1 \text{light}$

$\mu\{\text{flowers}|\text{light, time} = 1\} = (\beta_0 + \beta_2) + \beta_1 \text{light}$

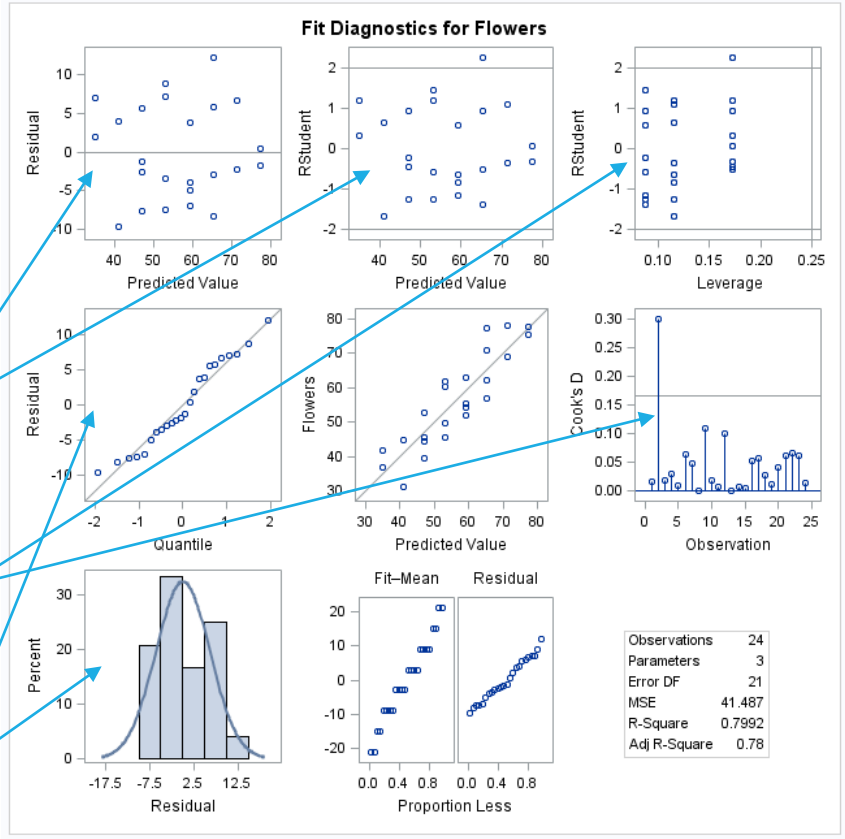
DISPLAY 9.3 The average number of flowers per plant versus the applied light intensity for each of the 12 experimental units in the meadowfoam study, with different plotting symbols for units with early (24 days prior to PFI) and late (at PFI) commencement of lighting regimen



Meadowfoam: Checking Assumptions

$\mu\{\text{flowers}|\text{light}\} = \beta_0 + \beta_1 light + \beta_2 time$

```
proc reg data = flower;
model flowers = intensity time;
run;
```



<u>Residuals:</u> No evidence of changing variance nor model misfit
(what does "Predicted Value" mean?)
<u>Influence:</u> No evidence of extreme or influential points
(High leverage + large residual)
<u>Normality:</u> No evidence of any issue

Meadowfoam: Conclusion

The REG Procedure
Model: MODEL1
Dependent Variable: Flowers

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3466.70045	1733.35023	41.78	<.0001
Error	21	871.23580	41.48742		
Corrected Total	23	4337.93625			

Root MSE	6.44107	R-Square	0.7992
Dependent Mean	56.13750	Adj R-Sq	0.7800
Coeff Var	11.47374		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	83.46417	3.27377	25.49	<.0001
Intensity	1	-0.04047	0.00513	-7.89	<.0001
Time	1	-12.15833	2.62956	-4.62	0.0001

We estimate that Increasing “Light Intensity” decreases the mean number of flowers per plant by 4.0 flowers per plant per 100 nano mol / m2 / sec. (95% confidence interval from 3.0 to 5.1)

Beginning the light treatment 24 days before PFI (time = 0) increased the mean number of flowers per plant by an estimated 12.2 flowers (95% confidence interval from 6.7 to 17.6)

We can draw the above causal relationship since the seeds were randomly assigned to each treatment group.