

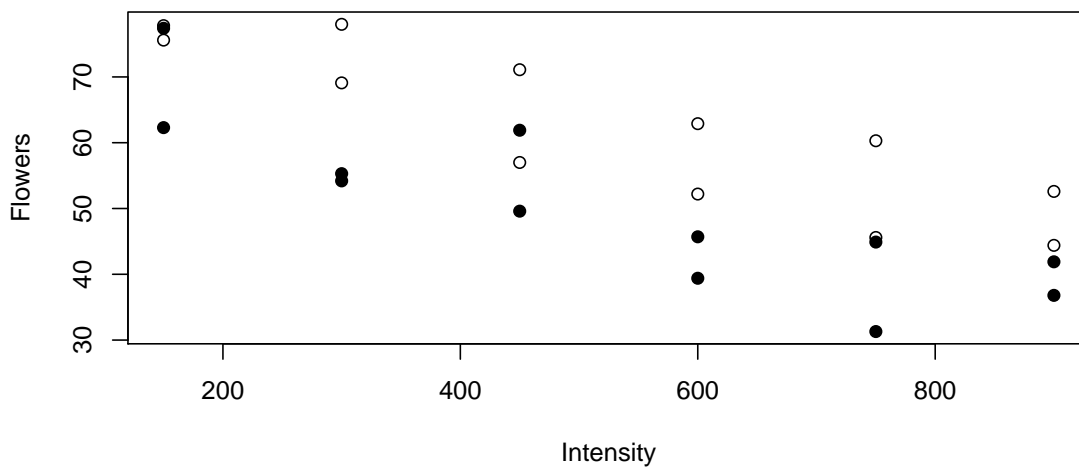
This document contains a summary of several analyses of the meadowfoam flowering data from Chapter 9, and provides examples of R code. The data is displayed on page 237 of your textbook and below. **Flowers** measures the average number of flowers per plant. **Time** is a categorical variable (or factor) **Intensity** is the light intensity ($\mu\text{mol}/\text{m}^2/\text{sec}$). for whether the light treatment begins early or late.

```
> case0901
```

	Flowers	Time	Intensity
1	62.3	late	150
2	77.4	late	150
3	55.3	late	300
4	54.2	late	300
5	49.6	late	450
6	61.9	late	450
7	39.4	late	600
8	45.7	late	600
9	31.3	late	750
10	44.9	late	750
11	36.8	late	900
12	41.9	late	900
13	77.8	early	150
14	75.6	early	150
15	69.1	early	300
16	78.0	early	300
17	57.0	early	450
18	71.1	early	450
19	62.9	early	600
20	52.2	early	600
21	60.3	early	750
22	45.6	early	750
23	52.6	early	900
24	44.4	early	900

The first step is to plot the data using two different symbols for the late and early points. The method here is to make an empty plot first (`type="n"`) and then to add the points (with `points`).

```
> early <- (Time == "early")
> late <- (Time == "late")
> plot(Intensity, Flowers, type = "n")
> points(Intensity[early], Flowers[early])
> points(Intensity[late], Flowers[late], pch = 16)
```



The first thing to try is to fit parallel lines, in other words, to assume no interaction between Intensity and Time. (We should also look at residual plots for each fit, but I am suppressing this for the moment.)

```
> fit1 <- lm(Flowers ~ Intensity + Time)
> summary(fit1)
```

Call:
lm(formula = Flowers ~ Intensity + Time)

Residuals:
Min 1Q Median 3Q Max
-9.652 -4.139 -1.558 5.632 12.165

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 83.464167 3.273772 25.495 < 2e-16 ***
Intensity -0.040471 0.005132 -7.886 1.04e-07 ***
Timelate -12.158333 2.629557 -4.624 0.000146 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.441 on 21 degrees of freedom
Multiple R-Squared: 0.7992, Adjusted R-squared: 0.78
F-statistic: 41.78 on 2 and 21 DF, p-value: 4.786e-08

```
> anova(fit1)
```

Analysis of Variance Table

Response: Flowers
Df Sum Sq Mean Sq F value Pr(>F)

```
Intensity 1 2579.75 2579.75 62.181 1.037e-07 ***
Time      1 886.95 886.95 21.379 0.0001464 ***
Residuals 21 871.24 41.49
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A second fit includes an interaction between Intensity and Time. This effectively fits two lines with different intercepts and slopes. In the formula, `Intensity * Time` does not stand for the product of intensity and time, but rather is expanded to mean `Intensity + Time + Intensity:Time`, or the two main effects plus an interaction term. The categorical variable Time is treated numerically as a variable that take the value 0 for one level (early, in this case) and the value 1 for the other level late.

```
> fit2 <- lm(Flowers ~ Intensity * Time)
> summary(fit2)
```

Call:

```
lm(formula = Flowers ~ Intensity * Time)
```

Residuals:

```
    Min      1Q  Median      3Q     Max
-9.516 -4.275 -1.422  5.473 11.938
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   83.146667   4.343305  19.144 2.49e-14 ***
Intensity     -0.039867   0.007435  -5.362 3.01e-05 ***
Timelate     -11.523333   6.142360  -1.876  0.0753 .
Intensity:Timelate -0.001210  0.010515  -0.115  0.9096
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.598 on 20 degrees of freedom

Multiple R-Squared: 0.7993, Adjusted R-squared: 0.7692

F-statistic: 26.55 on 3 and 20 DF, p-value: 3.549e-07

```
> anova(fit2)
```

Analysis of Variance Table

Response: Flowers

```
      Df Sum Sq Mean Sq F value    Pr(>F)
Intensity 1 2579.75 2579.75 59.2597 2.101e-07 ***
Time      1 886.95 886.95 20.3742 0.0002119 ***
Intensity:Time 1 0.58 0.58 0.0132 0.9095675
Residuals 20 870.66 43.53
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Up to this point, we have treated Intensity as a quantitative variable. We could also treat it as a factor with six levels. This would essentially be a two-way analysis of variance. Here is how to do this in R, first without an interaction term and then with one.

```
> fit3 <- lm(Flowers ~ as.factor(Intensity) + Time)
> summary(fit3)
```

Call:
lm(formula = Flowers ~ as.factor(Intensity) + Time)

Residuals:
Min 1Q Median 3Q Max
-8.979 -4.308 -1.342 5.204 10.204

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	79.354	3.629	21.869	6.92e-14 ***
as.factor(Intensity)300	-9.125	4.751	-1.921	0.071715 .
as.factor(Intensity)450	-13.375	4.751	-2.815	0.011919 *
as.factor(Intensity)600	-23.225	4.751	-4.888	0.000138 ***
as.factor(Intensity)750	-27.750	4.751	-5.841	1.97e-05 ***
as.factor(Intensity)900	-29.350	4.751	-6.178	1.01e-05 ***
Timelate	-12.158	2.743	-4.432	0.000365 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.719 on 17 degrees of freedom
Multiple R-Squared: 0.8231, Adjusted R-squared: 0.7606
F-statistic: 13.18 on 6 and 17 DF, p-value: 1.427e-05

```
> anova(fit3)
```

Analysis of Variance Table

Response: Flowers

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(Intensity)	5	2683.51	536.70	11.888	4.63e-05 ***
Time	1	886.95	886.95	19.646	0.0003649 ***
Residuals	17	767.47	45.15		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> fit4 <- lm(Flowers ~ as.factor(Intensity) * Time)
> summary(fit4)
```

Call:
lm(formula = Flowers ~ as.factor(Intensity) * Time)

Residuals:
Min 1Q Median 3Q Max
-7.550e+00 -4.675e+00 9.992e-16 4.675e+00 7.550e+00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	76.700	5.228	14.671	5.01e-09	***
as.factor(Intensity)300	-3.150	7.393	-0.426	0.67760	
as.factor(Intensity)450	-12.650	7.393	-1.711	0.11278	
as.factor(Intensity)600	-19.150	7.393	-2.590	0.02365	*
as.factor(Intensity)750	-23.750	7.393	-3.212	0.00746	**
as.factor(Intensity)900	-28.200	7.393	-3.814	0.00246	**
Timelate	-6.850	7.393	-0.927	0.37244	
as.factor(Intensity)300:Timelate	-11.950	10.456	-1.143	0.27536	
as.factor(Intensity)450:Timelate	-1.450	10.456	-0.139	0.89200	
as.factor(Intensity)600:Timelate	-8.150	10.456	-0.779	0.45080	
as.factor(Intensity)750:Timelate	-8.000	10.456	-0.765	0.45897	
as.factor(Intensity)900:Timelate	-2.300	10.456	-0.220	0.82959	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.393 on 12 degrees of freedom
 Multiple R-Squared: 0.8488, Adjusted R-squared: 0.7102
 F-statistic: 6.124 on 11 and 12 DF, p-value: 0.002028

> anova(fit4)

Analysis of Variance Table

Response: Flowers

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
as.factor(Intensity)	5	2683.51	536.70	9.8189	0.0006388	***
Time	1	886.95	886.95	16.2266	0.0016745	**
as.factor(Intensity):Time	5	111.55	22.31	0.4081	0.8341569	
Residuals	12	655.92	54.66			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1