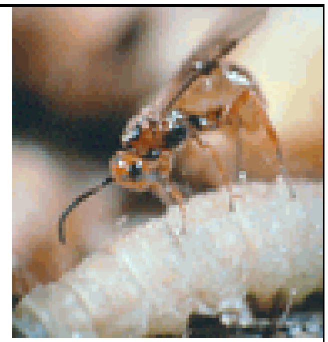
Data Set 3: Parasitoid Larval Competition on Different Species of Hosts

Statistical Setting

This Handout is an example of a "**general linear model**" analysis involving one **quantitative independent variable** and one **categorical ("qualitative") independent variable**, and a quantitative response variable. Both independent variables were of direct interest, so the analysis aims to determine how each affects the response and whether their effects interact. Following tests for these effects, confidence intervals for the pertinent parameters are computed. Standard residual plots also are presented, along with a lack-of-fit test.

Background and Data



B. hebetor female laying eggs on a meal moth caterpillar

As part of my dissertation research I conducted laboratory experiments to describe competition — *i.e.* the negative effects of increased density — among larvae of the parasitic wasp, *Bracon hebetor*, which parasitizes caterpillars of several species of moths. In the experiment described in this handout I investigated the effects of different species of hosts on the competition among the wasp larvae. I was most interested in whether the intensity of competition was related to the size of the host, hypothesizing that competition would be more intense when the wasp larvae were feeding on, and competing for, a smaller host.

I manipulated the density of *B. hebetor* eggs on individual hosts of each of three species, which for simplicity will be called "A", "B", and "C" in this handout. Species A is largest and C the smallest. Densities were 4, 8, 12, 16, and 20 eggs/host.

The response variable analyzed was adult wasp size, measured as the mean head width (distance between the eyes, in mm) of all emerging adults in each experimental unit (group of wasps on a host). There were 10 replicates per treatment (density x host-species combination), but in a few replicates no adult wasps emerged so there was no value for the mean size.

eggs, larva, pupa and adult *Plodia interpuntella* (Indian meal moth), a host of *B. hebetor*

Data Exploration

Graphical

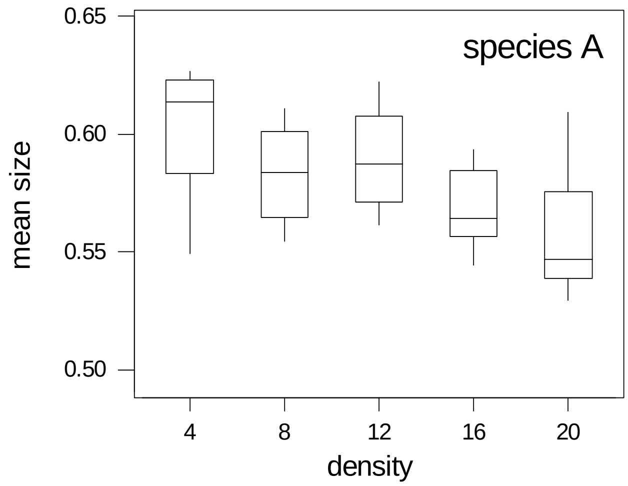
The boxplots on the following page show that wasp size decreased with increasing density on all three host species, and also that wasp size was similar on hosts of species A and B but smaller on species C. The boxplots also show that there is no general pattern of skew in the distribution, or great differences in spread. One outlier is indicated, for 4 eggs on species B, but this appears to be caused by the small spread of the other observations in this group.

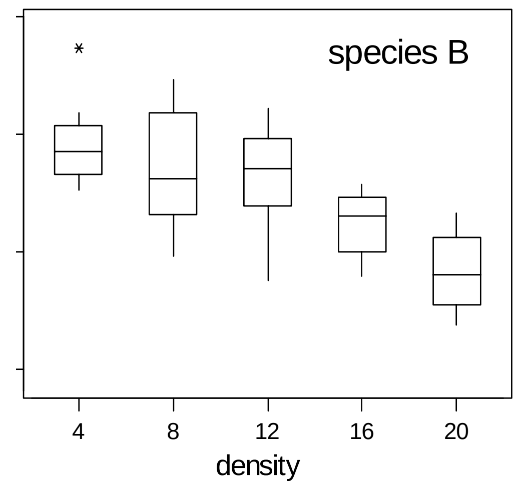
Comparison among host species in the trends of wasp size in relation to density is clearer in the plot at the bottom of the next page, showing the LOWESS of this relationship for each host species. At the lowest densities wasps are slightly larger on species A than on B or C, but the difference is slight. As density increases, differences among the host species also increase, with wasps on species C becoming considerably smaller than those on A or B. In other words, the decrease in size with increasing density is steepest on species C and quite similar between species A and B.

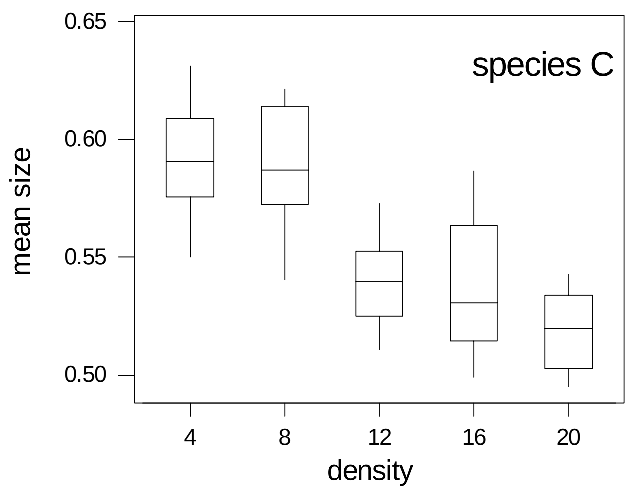
Summary Statistics

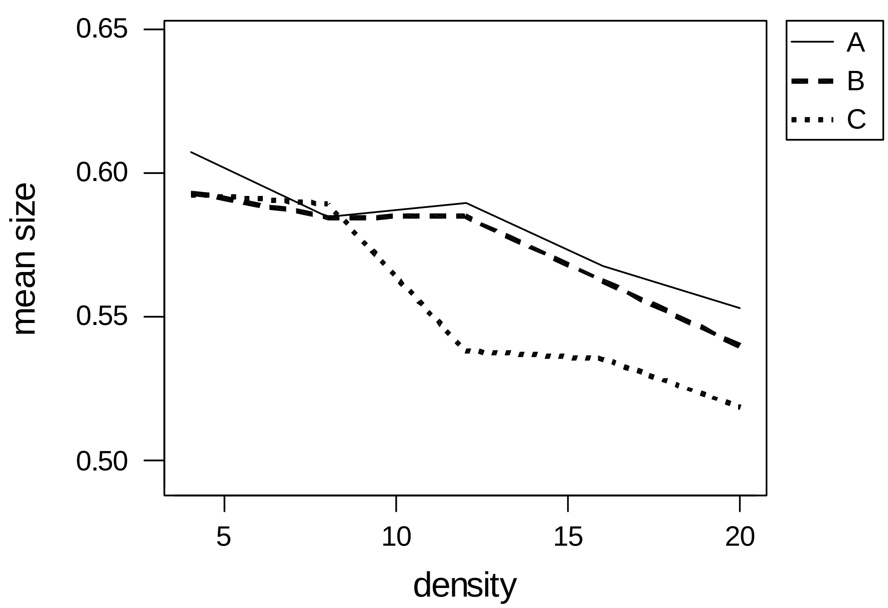
		n	.mean	.median	.std dev
A	4	9	0.60243	0.61383	0.02621
	8	10	0.58416	0.58391	0.01944
	12	10	0.58985	0.58738	0.02068
	16	10	0.56833	0.56422	0.01640
	20	10	0.55807	0.54672	0.02741
B	4	10	0.59568	0.59240	0.01797
	8	9	0.58615	0.58106	0.02526
	12	10	0.58178	0.58524	0.02177
	16	9	0.56185	0.56494	0.01342
	20	10	0.54055	0.53988	0.01667
C	4	10	0.59224	0.59085	0.02286
	8	10	0.58830	0.58706	0.02539
	12	9	0.53853	0.53936	0.01883
	16	10	0.53714	0.53068	0.02778
	20	8	0.51850	0.51950	0.01688

These statistics quantify the pattern seen in the graphs, that in general the mean wasp size decreased with increasing egg density, and was largest on host species A and smallest on species C. These statistics also show that the variability among replicates was reasonably similar for all experimental conditions, and differences in variability were not consistently related to the independent variables.


species A
mean size
0.65
0.60
0.55
0.50
4
8
12
16
20
density


species B
*
4
8
12
16
20
density


species C
mean size
0.65
0.60
0.55
0.50
4
8
12
16
20
density


A
B
C
mean size
0.65
0.60
0.55
0.50
5
10
15
20
density

Inference

Model

The categorical variable "host species" can be modeled using two indicator variables, *e.g.* `hsp_A`, coded as 1 for species A and 0 otherwise; and `hsp_B`, coded as 1 for species B and 0 otherwise (this is the coding used by Minitab). The host species x density interaction would then be modeled by multiplying each of these indicator variables by density, giving the variables `intxn_A` and `intxn_B`. The full model, with this coding of the indicator variables, would be:

$$Y_i = \beta_0 + \beta_1 \text{density}_i + \beta_2 \text{hsp\_A}_i + \beta_3 \text{hsp\_B}_i + \beta_4 \text{intxn\_A}_i + \beta_5 \text{intxn\_B}_i + \varepsilon_i$$

For the different host species the models would be:

A: $Y_i = \beta_0 + \beta_2 + (\beta_1 + \beta_4)\text{density}_i + \varepsilon_i$

B: $Y_i = \beta_0 + \beta_3 + (\beta_1 + \beta_5)\text{density}_i + \varepsilon_i$

C: $Y_i = \beta_0 + \beta_1 \text{density}_i + \varepsilon_i$

Hypothesis tests

ANOVA table for model:

SOURCE	DF	SS	MS	F	P
Model	5	0.078547	0.015709	32.09	0.000
Error	138	0.067559	0.000490		
Total	143	0.146106			

We conclude from this very significant overall *F*-test that there is strong evidence that at least some of β_1–β_5 are not 0, *i.e*. that one or both of the experimental factors had an effect on mean wasp size.

Coincidence of Regressions:

The test of whether the regressions of size against density on the three host species are coincident — have the same slopes and intercepts — tests the following hypotheses:

H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ vs. H_a: not all of $\beta_2 \ldots \beta_5 = 0$

These hypotheses could be tested by comparing the *SSE* for the reduced model (*i.e*. the simple linear regression of mean size on density) to that for the full model (above). Alternatively, added-in-order *SS*s can be used, so long as density is entered first. If a regression procedure is used with indicator variables coded as above, the added-in-order *SS*s are:

density	0.058263
hsp_A	**0.008506**
hsp_B	**0.007322**
intxn_A	**0.002446**
intxn_B	**0.002011**

The added *SS* for β_2 through β_5 (*i.e.* for the species and species x density effects) is the sum of the last four of these added-in-order *SS*s, equal to 0.020285, with 4 degrees of freedom.

If instead a GLM procedure is used the output would (or could be gotten to) include the following or equivalent::

Source	DF	Seq SS
density	1	0.058263
hostsp	2	**0.015828**
hostsp*density	2	**0.004457**

Here the relevant terms are those for the species ('`hostsp`') and species x density effects ('`hostsp*density`'), which sum to the same SS and df as above.

The multiple-partial *F* test for the hypothesis of coincident regressions then is

SOURCE	DF	SS	MS	F	P
Added terms	4	0.020285	0.005071	10.36	0.000
Error	138	0.067559	0.000490		

We therefore conclude there is strong evidence the regression relationship is not the same for all three host species, and proceed to test whether the regressions are parallel.

Parallelism of Regressions (Equality of Slopes):

For this test the hypotheses are:

H_0: $\beta_4 = \beta_5 = 0$
H_a: not both of β_4 and $\beta_5 = 0$

The added *SS* for this test is the sum of the last two of the added-in-order *SSe*s from the indicator variable–regression output above, or simply the last term of the GLM output.. The multiple-partial *F* test is

SOURCE	DF	SS	MS	F	P
Added terms	2	0.004457	0.002228	4.55	0.012
Error	138	0.067559	0.000490		

We therefore conclude there is strong evidence the regression slopes are not the same for all three host species.

Since the regressions are not parallel, I don't feel it would be sensible to test further hypotheses about either differences in the intercepts for the different host species (*i.e.* β_2 and β_3) or the slope for host species C (*i.e.* β_1).

Tests for individual parameters:

Predictor	.Coef	St. dev	*t*	P
Constant	0.615134	0.007470	82.35	0.000
density	-0.0049995	0.0005778	-8.65	0.000
hsp_A	-0.00328	0.01063	-0.31	0.758
hsp_B	-0.00152	0.01053	-0.14	0.885
intxn_A	0.0023908	0.0008080	2.96	0.004
intxn_B	0.0016290	0.0008038	2.03	0.045

These results indicate that the slopes for species A and B both are significantly different from that for species C (these differences in slope being modeled as β_4 and β_5, the estimates of which are the terms labelled `intxn_B` and `intxn_B` in the output). The `density` term also is highly significantly different from 0; this is the slope on species C.

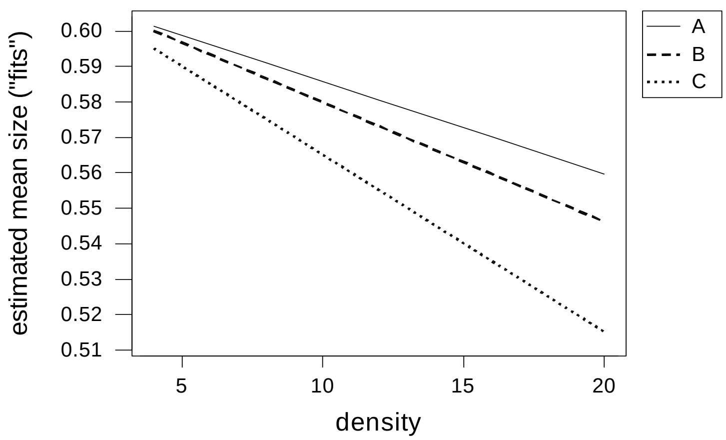
Estimation

<u>The models:</u>

The individual parameter estimates in the preceding output produce the following models:

```
A:   estimated meansize = 0.6118 - 0.00261 density
B:   estimated meansize = 0.6136 - 0.00337 density
C:   estimated meansize = 0.6151 - 0.00500 density
```

These three fitted lines are shown on the plot below. As the plot shows, the fitted model matches several patterns noted in the plot of smoothers for the three relationships, shown in the



Data Exploration section above: size decreases with density, on all host species; wasps tend to be smaller on species C and larger on species A, for a given density; and the slope — the effect of density — is steepest on the small species C and least steep on the large species A.

<u>Confidence intervals for slopes:</u>

Since the slopes for host species A and B are the sums of two coefficients, their standard errors must be calculated using the estimated variance-covariance matrix of the coefficients, which is *MSE* $(\mathbf{X}'\mathbf{X})^{-1}$. *MSE* = 0.00049 (from the ANOVA table); $(\mathbf{X}'\mathbf{X})^{-1}$ is

	$\mathbf{b_0}$	$\mathbf{b_1}$	$\mathbf{b_2}$	$\mathbf{b_3}$	$\mathbf{b_4}$	$\mathbf{b_5}$
$\mathbf{b_0}$	0.113974	-0.007950	-0.113974	-0.113974	0.007950	0.007950
$\mathbf{b_1}$		0.000682	0.007950	0.007950	-0.000682	-0.000682
$\mathbf{b_2}$			-0.000682	0.113974	-0.015876	-0.007950
$\mathbf{b_3}$				0.226644	-0.007950	-0.015603
$\mathbf{b_4}$					0.001333	0.000682
$\mathbf{b_5}$						0.001320

Since the estimated slope for host species A is $b_1 + b_4$, its standard error is

$$s\{b_1 + b_4\} = \sqrt{s^2\{b_1\} + s^2\{b_4\} + 2s\{b_1, b_4\}}$$
$$= \sqrt{0.00049(0.000682 + 0.001333 + 2 \cdot -0.000682)} = 0.000565$$

In the same way, s\{slope for species B\} = $s\{b_1 + b_5\}$ = 0.000559. Since the estimated slope for species C is simply b_1, its standard error is as given in the table of parameter estimates: 0.0005778.

Confidence intervals are then constructed in the usual way with critical values from the appropriate t distribution (here, df = 138). For instance, for 98% individual confidence intervals (giving Bonferroni-adjusted overall confidence of 94%), the critical value $t_{0.99;\,138} = 2.3537$ would be used, producing the following confidence intervals:

```
A:    -0.00261 ± 2.3537x0.000565 = (-0.00394, -0.00128)
B:    -0.00337 ± 2.3537x0.000559 = (-0.00469, -0.00205)
C:    -0.00500 ± 2.3537x0.000578 = (-0.00636, -0.00364)
```

Comparisons Between Slopes:

Just as multiple comparisons among group means often are useful following a significant overall ANOVA result, here comparisons among the three slopes are useful to determine which of the slopes differ from each other. These comparisons can take the form of t-tests or confidence intervals; I will show both. In this case, since the host species decreased in size from species A through species C, and it was expected that the intensity of competition would be greater on smaller hosts, one might do one-sided test, but since I was not really certain in advance this expectation would be met, and would want to detect differences in the other direction if present, I have used two-sided tests.

Here the comparisons of host species C with each of A and B are easy, since those differences are estimated by b_4 and b_5 respectively, and their standard errors and t-tests are given in the table of estimates above. The slope difference A – B is estimated by $(b_4 - b_5)$ = 0.00076; its standard error (calculated as in the previous section) is 0.000795. Tests and simultaneous 96% confidence intervals for these three differences are:

	estimate	**se**	**t**	**P**	**CI**
A - B	0.000762	0.000795	0.958	0.830	-0.001109, 0.002633
A - C	0.002391	0.000808	2.959	0.004	0.000489, 0.004296
B - C	0.001629	0.000804	2.027	0.045	-0.000263, 0.003521

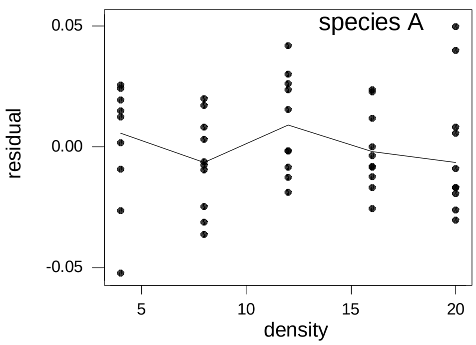
From this we can conclude that there is no evidence the slopes for species A and B differ; one might therefore think about fitting a model combining these two host species, but I did not feel this was worth doing. The slopes for species A and C do clearly differ, while the difference in slopes between B and C is of marginal significance.

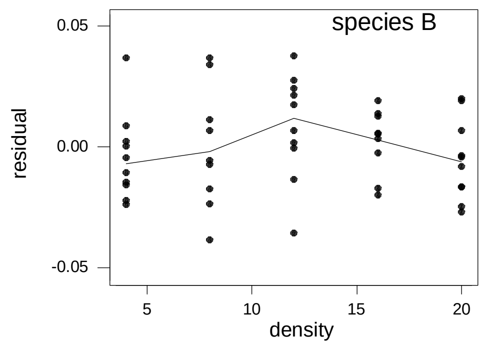
Diagnostics

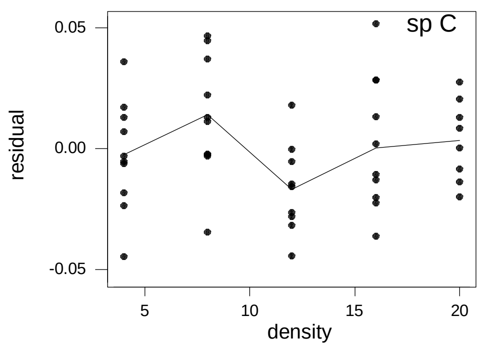
Residual Plots

<u>Residuals *vs* variables:</u>

In this case, with only one quantitative independent variable, separate plots can be done for each level of the categorical independent variable (*i.e.* each host species); on the next page I show plots of residuals *vs*. the independent variable but plots of residuals *vs*. fits could also be used (and would be mirror images of the plots shown here.)



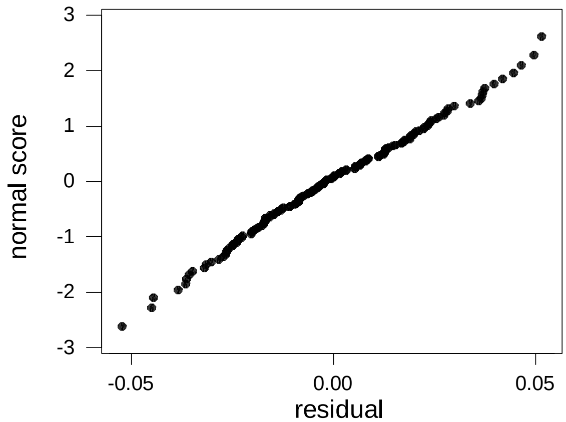




The plots of residuals *vs.* density show a hint of nonlinearity on species A and B: residuals tend to be larger at intermediate densities. This pattern is so slight, however, that I do not feel it needs any remedial action. (One simple remedy would be to fit a quadratic model, but in fact any higher-order polynomial terms are not statistically significant.)

The other concern in these plots is that there appears to be slightly less variation at high densities than low, for species B and C. The standard deviations shown earlier (with descriptive statistics), however, do not show any clear trend. So again I see no need for remedial action.

Distribution of residuals:



The residuals obviously are quite close to normally distributed.

Lack-of-fit test

SOURCE	DF	SS	MS	F	P
Model	5	0.078547			
Lack-of-Fit	9	0.007131	0.000792	1.692	0.0973
Pure Error	129	0.060428	0.000468		
Total	143	0.146106			

This result does not give clear evidence that the model is inadequate, though it also certainly does not give strong support that the model structure is indeed correct.

Conclusions from Diagnostics

The only concern from these diagnostics is the slight suggestion of nonlinearity. In my this is not substantial enough to invalidate the general conclusions of this analysis. There are no other evident problems with the model.

Conclusions

Taking the linear regressions as valid descriptions of the larval competition — the decrease in size with increasing egg density — on the three host species, I conclude that:

(1) There was larval competition, on all host species: Increased egg density caused a decrease in the mean size of surviving wasps.

(2) There was little if any effect of host species in the absence of competition: Wasp sizes were similar across the three host types a low density (*i.e.* the intercepts were similar).

(3) The intensity of larval competition differed among the host species, being strongest on the smallest species (C) and weakest on the largest species (A): The regression slopes were significantly different (interaction was statistically significant), with the slopes for both species A and B being significantly flatter than for species C.