Data Set 9: Laysan Finch Beak Widths – Island, Sex and Year Effects

Statistical Setting and Background

This handout describes several versions of an example of a three-factor ANOVA. First I show a fairly complete analysis of a balanced (*i.e.* equal replication) subset of the data, treating all three factors as fixed effects. I then show how parts of this analysis are modified if one of the factors is random. Finally, I show some of the complications arising in analysis of the full data set, which is very unbalanced.

The data are from Sheila Conant and Marie Morin's study of Laysan Finch beak morphology. In this handout beak width will be analyzed, to determine whether and how it varies between populations (Laysan and North islands), between sexes, and among years (1985-87). Only data from mature adults will be used.

Balanced Data Set

Data

The actual data set is quite large and very unbalanced, so I produced a balanced subset of the data by randomly selecting 8 observations for each of the 12 combinations of island, sex and year. This or the full data set can be sent to anyone requesting a copy.

Preliminary Data Exploration

Descriptive statistics:

			n	Mean	StDev	Median
Laysan	male	1985	8	0.846	0.0207	0.845
		1986	8	0.800	0.0227	0.805
		1987	8	0.824	0.0269	0.832
	female	1985	8	0.775	0.0233	0.780
		1986	8	0.784	0.0233	0.785
		1987	8	0.752	0.0292	0.754
North	male	1985	8	0.823	0.0191	0.830
		1986	8	0.800	0.0333	0.819
		1987	8	0.832	0.0189	0.834
	female	1985	8	0.769	0.0160	0.770
		1986	8	0.782	0.0212	0.777
		1987	8	0.796	0.0230	0.790

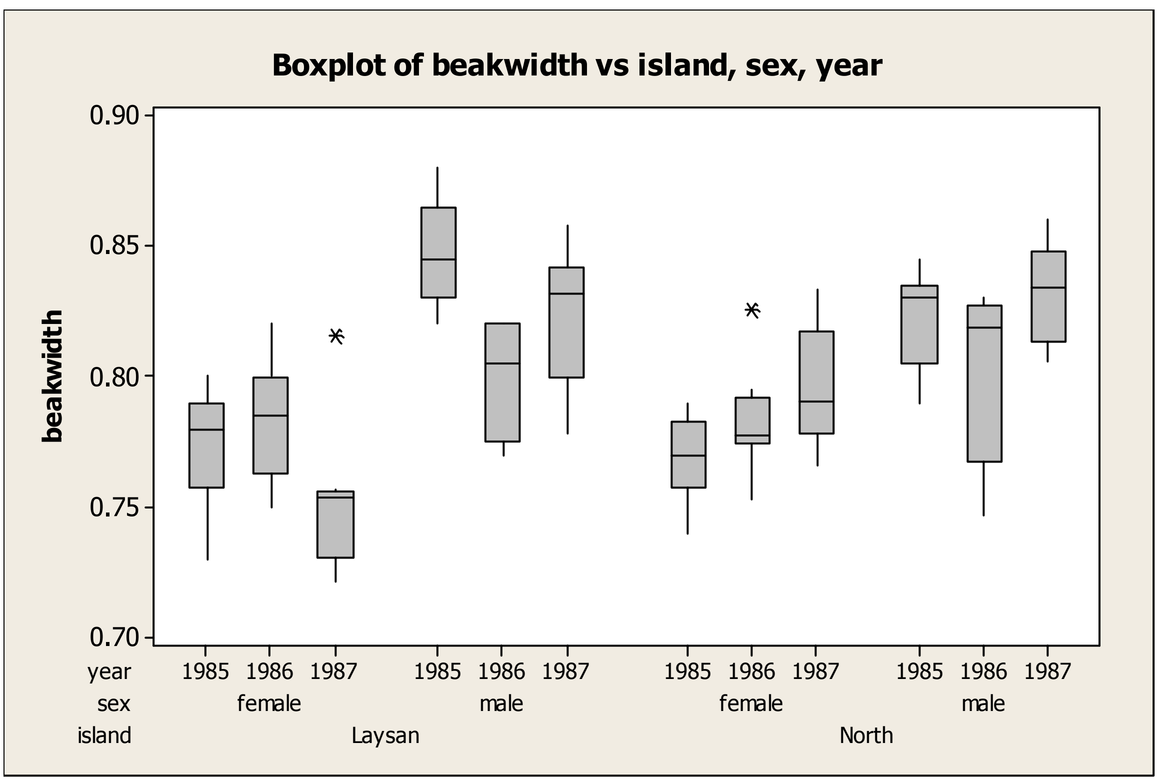
Descriptive statistics (above and on the next page) show that females tend to have narrower beaks than do males, on both islands and in all years. There is little if any evidence of a difference between islands, within a sex and/or year. There also are no big differences among years. There also are suggestions of all three two-way interactions: the trend over the years differs between sexes, and also between islands, and the difference between sexes is larger on Laysan than North island.

		n	Mean	StDev	Median
Laysan	1985	16	0.811	0.0425	0.810
	1986	16	0.792	0.0237	0.790
	1987	16	0.788	0.0462	0.785
North	1985	16	0.796	0.0328	0.790
	1986	16	0.791	0.0285	0.781
	1987	16	0.814	0.0275	0.815
male	1985	16	0.835	0.0226	0.835
	1986	16	0.800	0.0275	0.819
	1987	16	0.828	0.0228	0.834
female	1985	16	0.772	0.0196	0.777
	1986	16	0.783	0.0215	0.779
	1987	16	0.774	0.0340	0.772

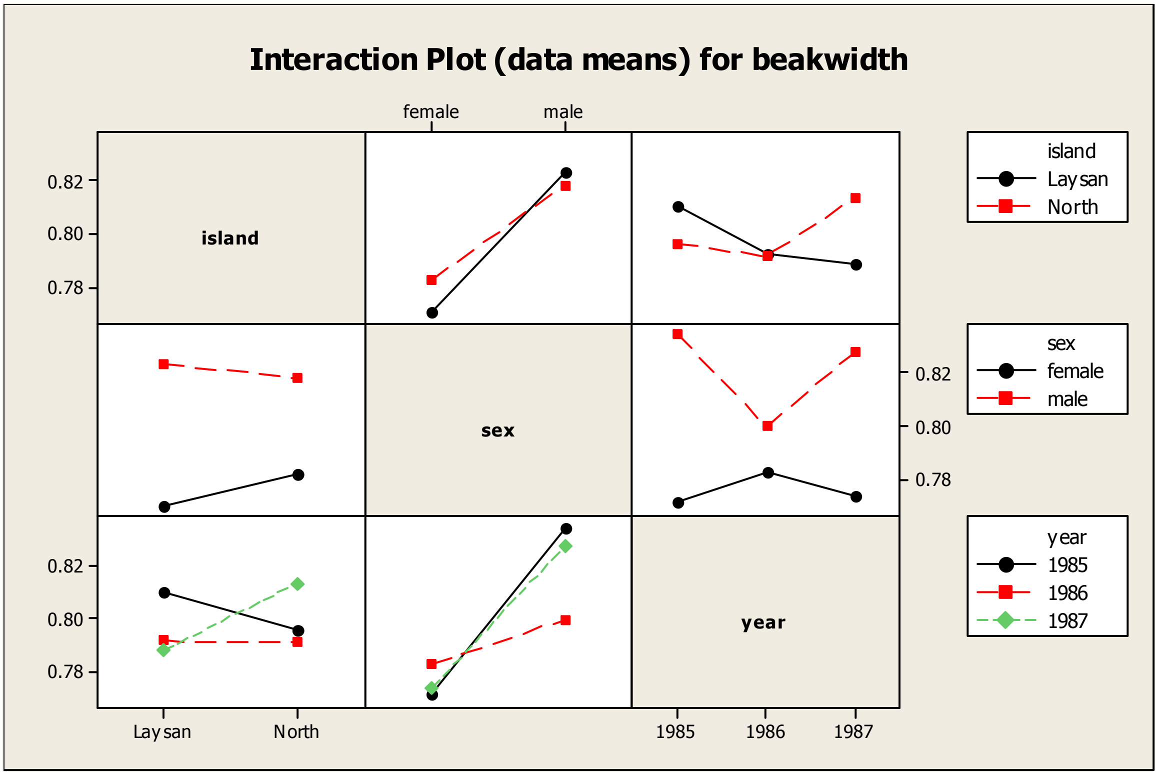
		n	Mean	StDev	Median
Laysan	male	24	0.82	0.030	0.82
	female	24	0.77	0.028	0.77
North	male	24	0.82	0.027	0.82
	female	24	0.78	0.022	0.78
Laysan		48	0.797	0.0392	0.791
North		48	0.800	0.0307	0.797
male		48	0.821	0.0283	0.823
female		48	0.776	0.0257	0.778
1985		32	0.803	0.0381	0.795
1986		32	0.792	0.0258	0.790

Boxplots

The boxplots show the same patterns noted in the descriptive statistics. The boxplots also show there are two somewhat anomalous observations (females with wide beaks, on Laysan in 1987 and North Island in 1986). Males on North Island in 1986 were more variable than the other samples, while North Island females in 1985 were somewhat less variable.

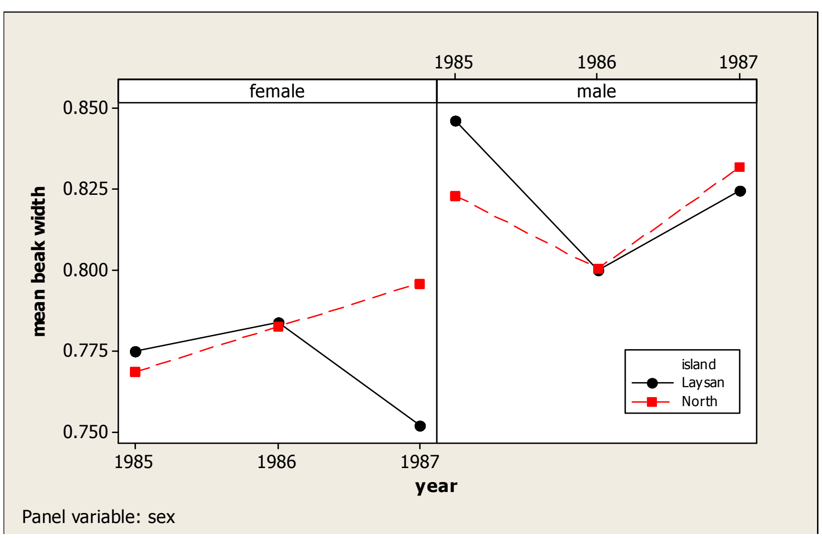


Plots of Sample Means

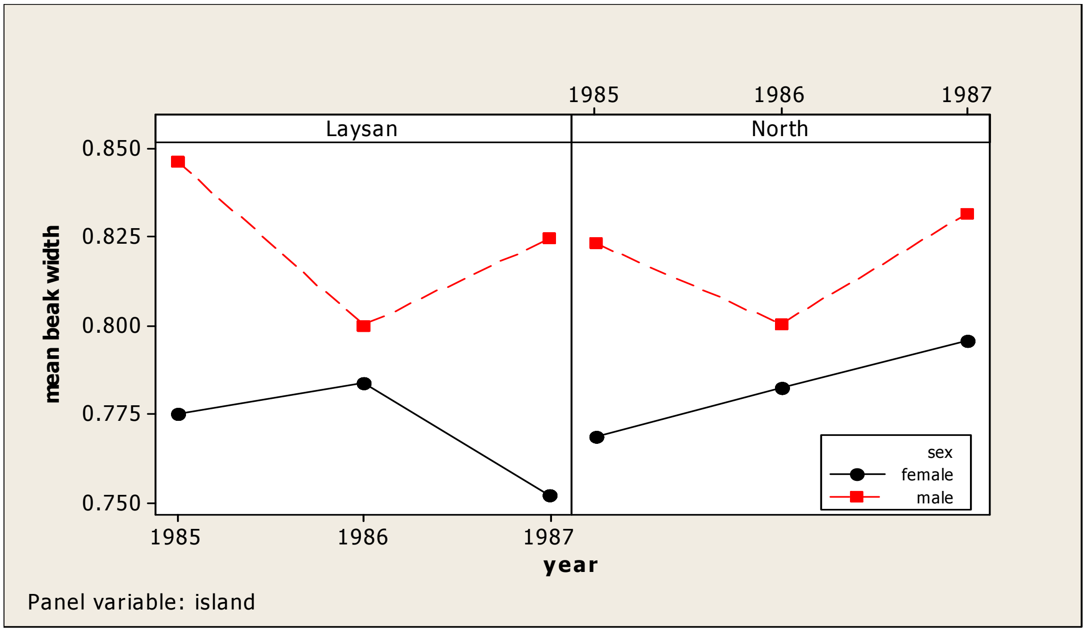


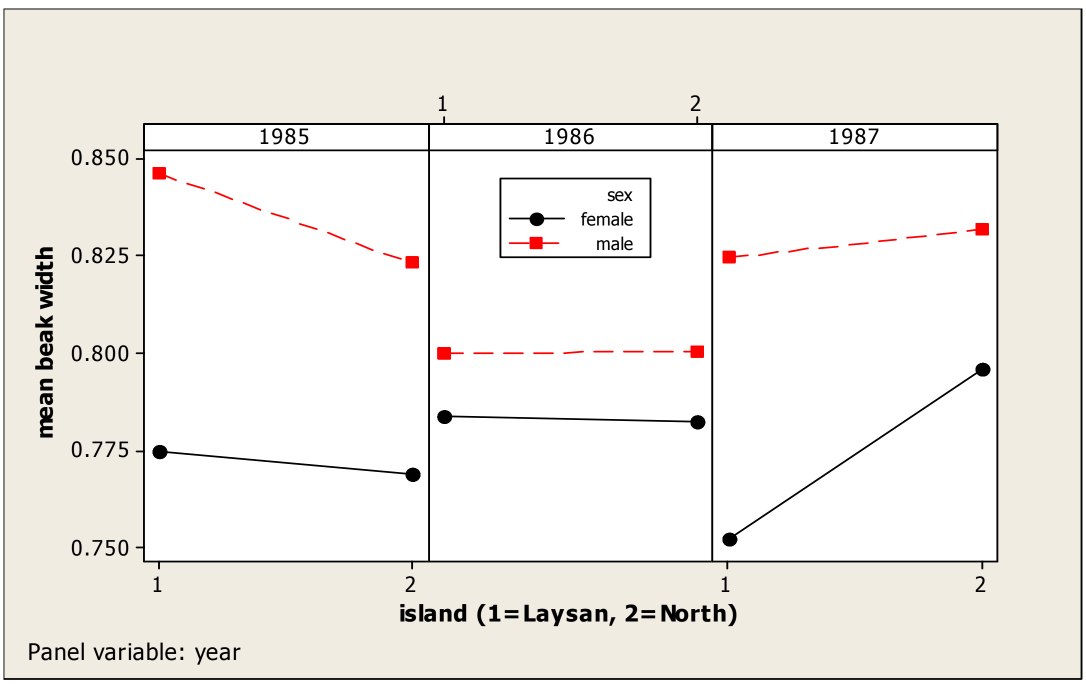
The difference between the sexes, and the absence of consistent differences between islands or among years, are evident in the interaction plots above. These plots also more clearly show the interactions between sex and year, and between island and year. The difference between sexes was considerably smaller in 1986 than in the other years. In 1985 Laysan birds had slightly wider beaks than did North Island birds, in 1986 there was very little difference, and in 1987 the North Island birds had wider beaks.

Interaction plots of two factors, separated by levels of the third factor, are necessary to examine the possibility of a three-way interaction. These plots, in three



arrangements, are shown here and on the next page. The most striking feature of these plots is that while beak widths of males on both islands and females on North island all were greater in 1987 than 1986, those for females on Laysan were less in 1987. Apart from this, there is little indication of a three-way interaction.





Analysis

ANOVA

Source	DF	SS	MS	F	P
isl	1	0.0002836	0.0002836	0.51	0.477
sex	1	0.0479273	0.0479273	86.33	**0.000**
year	2	0.0024518	0.0012259	2.21	0.116
isl*sex	1	0.0017768	0.0017768	3.20	0.077
isl*year	2	0.0066462	0.0033231	5.99	**0.004**
sex*year	2	0.0094182	0.0047091	8.48	**0.000**
isl*sex*year	2	0.0014634	0.0007317	1.32	0.273
Error	84	0.0466344	0.0005552		
Total	95	0.1166017			

The results of the formal analysis of the data are consistent with the patterns seen in the plots above. The three-way interaction is not significant. The two-way interactions of year with each of the other factors, however, are highly significant, and the interaction of sex and island also has a fairly small *P*-value. Given these interactions, caution is required in interpreting main effects. For these data, though, I believe the highly significant main effect of sex is meaningful: males' beaks are substantially wider than females' beaks in all years, on both islands.

Note that the *P*-values given in the computer output are for each test individually. The Kimball inequality can be used to control the significance level for the family of seven tests. For instance, for maintain a combined significance of $\alpha = 0.05$, each test would be conducted at $\alpha^* = 1 - (0.95)^{1/7} = 0.0073$. The three effects which are clearly significant in individual tests remain so at this family level of significance.

Analysis of Effects

Given the presence of several interactions, further analysis should be in terms of treatment (cell) means. Since the three-way interaction was not significant, one approach would be to analyze one or more of the two-way interactions averaging across the third factor (*e.g.* compare sexes, for each island, combining years). However, since there is strong evidence for two of the two way interactions, and a suggestion of the third, these probably would not be useful comparisons (*e.g.* the sex x year interaction makes it questionable what a comparison of sexes averaging over years would mean).

Probably the best that can be done in a situation such as this is to make appropriate pairwise comparisons among treatment means: compare the sexes for each island x year combination, compare the islands for each sex x year combination, and compare the years for each sex x island combination. This leads to a total of 24 comparisons, out of the 66 possible among the 12 treatment means; the 42 others are uninteresting (*e.g.* Laysan males in 1985 *vs*. North Island females in 1987). I conducted these comparisons (next page) at a Bonferroni-corrected family significance level of $\alpha = 0.10$; specifically, I used Fisher multiple comparisons in a one-way (12 groups) ANOVA in Minitab with the testwise α set at 0.10/24 = 0.00417, and then ignored the uninteresting comparisons.

Comparisons between the sexes show that males did have wider beaks, on both islands, in 1985 and 1987, but the sex differences were not significant in 1986:

Comparison	Lower	Center	Upper
Laysan Female 1985 - Laysan Male 1985	**0.03885**	**0.07125**	**0.10365**
" " 1986 - " " 1986	-0.01615	0.01625	0.04865
" " 1987 - " " 1987	**0.03998**	**0.07238**	**0.10477**
North Female 1985 - North Male 1985	**0.02198**	**0.05437**	**0.08677**
" " 1986 - " " 1986	-0.01440	0.01800	0.05040
" " 1987 - " " 1987	**0.00348**	**0.03587**	**0.06827**

In contrast, only one of the 12 between-year comparisons was significant (1985 *vs.* 1986, in Laysan males):

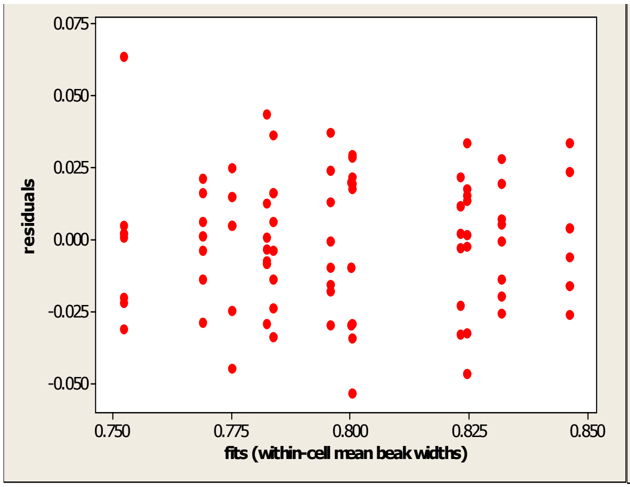
Comparison	Lower	Center	Upper
Laysan Female 1985 - Laysan Female 1986	-0.02365	0.00875	0.04115
" " " - " " 1987	-0.05527	-0.02288	0.00952
" " 1986 - " " 1987	-0.06402	-0.03163	0.00077
North " 1985 - North " 1986	-0.01877	0.01363	0.04602
" " " - " " 1987	-0.00527	0.02712	0.05952
" " 1986 - " " 1987	-0.01890	0.01350	0.04590
Laysan Male 1985 - Laysan Male 1986	**-0.07865**	**-0.04625**	**-0.01385**
" " " - " " 1987	-0.05415	-0.02175	0.01065
" " 1986 - " " 1987	-0.00790	0.02450	0.05691
North " 1985 - North " 1986	-0.05515	-0.02275	0.00965
" " " - " " 1987	-0.02377	0.00862	0.04102
" " 1986 - " " 1987	-0.00102	0.03138	0.06377

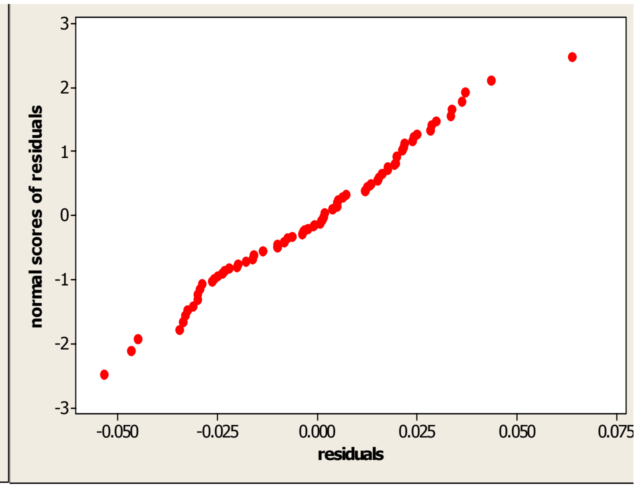
. Similarly, only one of the six between-island comparisons was significant (North Island females had wider beaks than Laysan females in 1987):

Comparison	Lower	Center	Upper
Laysan Female 1985 - North Female 1985	-0.03865	-0.00625	0.02615
" " 1986 - " " 1986	-0.03377	-0.00138	0.03102
" " 1987 - " " 1987	**0.01135**	**0.04375**	**0.07615**
Laysan Male 1985 - North Male 1985	-0.05552	-0.02312	0.00927
" " 1986 - " " 1986	-0.03202	0.00037	0.03277
" " 1987 - " " 1987	-0.02515	0.00725	0.03965

Assessment of "Aptness" of ANOVA Model

One problem applying ANOVA to these data has already been mentioned: although most of the sample standard deviations are quite similar, one was somewhat small and one somewhat large, giving a maximum:minimum ratio of slightly over 2. However, it must be realized that this is for a total of 12 samples, and that the more samples one has, the greater the variation in sample standard deviations even if all true variances are identical. Also, the samples sizes are equal, which makes the analysis more robust to unequal variance. Finally, a plot of residuals *vs.* fitted values (sample means) showed no pattern. I therefore see no reason to be concerned about non-homogeneous variances.





Another important issue to assess is normality. The normal quantile-quantile plot of all 96 residuals (deviations from group means) showed the data to be quite close to a normal distribution, except for one very minor high outlier.

I therefore see no reason not to accept the ANOVA analysis as valid.

Conclusions

The only consistent and substantial difference is between the sexes, with males having wider beaks. Even this difference, though, varied substantially among years. Similarly, differences between the islands were inconsistent, to the extent that in half of the comparisons between islands, within a sex and year, Laysan beaks were wider, and in the other half of comparisons North island beaks were wider.

To me (admittedly not a bird ecologist), the marked differences between consecutive years are surprising, and would deserve further investigation.

Balanced Data Set — Year as a Random Effect

The three consecutive years 1985 - 1987 clearly were not a random sample from some "population" of years. It might nonetheless be reasonable to consider these years as effectively a random sample of yearly environmental conditions (assuming that there is little correlation from year to year in the important environmental variables), and thus to treat the year effect as random. This would permit generalization beyond the three years actually studied. (Sex is a fixed effect. Island also is a fixed effect, since these were two of the three islands in the entire universe with substantial populations of Laysan Finch.)

Treating year as a random effect does not change the MSes in the ANOVA table, but does change their expected values (EMSes) and thus the appropriate tests; these depend on whether the restricted or unrestricted model is adopted. In this case the restricted model might be preferred due to the possibility of competitive interactions (and thus negative correlations of beak widths) between male and female birds on each island.

Preliminary explorations and assessment of model "aptness" in this mixed model would be the same as for the fixed-effects model above, and will not be repeated here.

Unrestricted Model

EMSes

For the unrestricted model (assuming no negative correlations among sexes or islands, within years), the EMSes are:

```
Source           Variance Error Expected Mean Square
                component  term (using unrestricted model)
  1 isl                     5   (8) + 8(7) + 16(5) + Q[1,4]
  2 sex                     6   (8) + 8(7) + 16(6) + Q[2,4]
  3 year          -0.00019  *   (8) + 8(7) + 16(6) + 16(5) + 32(3)
  4 isl*sex                 7   (8) + 8(7) + Q[4]
  5 isl*year       0.00016  7   (8) + 8(7) + 16(5)
  6 sex*year       0.00025  7   (8) + 8(7) + 16(6)
  7 isl*sex*year   0.00002  8   (8) + 8(7)
  8 Error          0.00056      (8)
```

From these EMSes it can be seen that each fixed-effect main effect is to be tested over its interaction with the random effect (years). All three two-way interactions are to be tested over the three-way interaction, and the latter is the only effect to be tested against MSE. No exact test is possible for the year main effect, but a *quasi-F* test can be constructed as

$$F^{**} = \frac{MS[year]}{MS[isl*year] + MS[sex*year] - MS[isl*sex*year]} \quad .$$

ANOVA

Source	DF	SS	MS	F	P
isl	1	0.0002836	0.0002836	0.09	0.798
sex	1	0.0479273	0.0479273	10.18	0.086
year	2	0.0024518	0.0012259	*	
isl*sex	1	0.0017768	0.0017768	2.43	0.259
isl*year	2	0.0066462	0.0033231	4.54	0.180
sex*year	2	0.0094182	0.0047091	6.44	0.134
isl*sex*year	2	0.0014634	0.0007317	1.32	0.273
Error	84	0.0466344	0.0005552		

The quasi-F for testing the year main effect is

$$F^{**} = \frac{0.0012259}{0.0033231 + 0.0047091 - 0.0007317} = 0.168 .$$

Minitab gives an approximate *df* for the denominator of 3.16:

	Source	Error DF	Error MS	Synthesis of Error MS
3	year	**3.16**	0.0073005	(5) + (6) - (7)

With an *F* statistic much smaller than 1, there is no real need to calculate a *P*-value. The result given by Minitab is

Source	DF	SS	MS	F	P	
year	2	0.0024518	0.0012259	0.17	0.853	x

The results of this analysis are quite different than for the fixed-effects model: there is a suggestion of a difference between the sexes, but no other effects are close to being significant. It is worth nothing, though, that the *F** statistics for several effects are considerably larger than 1; their lack of significance results from having very few denominator degrees of freedom.

Analysis of Effects

If we take the hypothesis tests at face value, no further analyses are warranted since no effects are significant. At most, we might compute confidence intervals for the male and female overall means, or for their difference. For the latter, *MS*(sex x year) and its *df* would be used in place of *MSE* in the usual *t* interval; the 90% CI is

$$(\overline{Y}_{M..} - \overline{Y}_{F..}) \pm t[0.95;2]\sqrt{\frac{MS[\text{sex x year}]}{48}(1^2 + (-1)^2)}$$

$$= (0.821 - 0.7763) \pm 2.92\sqrt{0.0047/24}$$

$$= 0.0447 \pm 0.04086 = (0.0038, 0.0856)$$

Restricted Model

EMSes

Source	Variance component	Error term	Expected Mean Square (using restricted model)
1 isl		5	(8) + 16(5) + 48Q[1]
2 sex		6	(8) + 16(6) + 48Q[2]
3 year	0.00002	8	(8) + 32(3)
4 isl*sex		7	(8) + 8(7) + 24Q[4]
5 isl*year	0.00017	8	(8) + 16(5)
6 sex*year	0.00026	8	(8) + 16(6)
7 isl*sex*year	0.00002	8	(8) + 8(7)
8 Error	0.00056		(8)

The tests for effects involving only fixed effects (*i.e.* island or sex main effects or their interaction) are the same as in the restricted model. All effects involving the random factor year, however, now are tested over MSE; note that an exact test now is possible for the year main effect.

ANOVA

Source	DF	SS	MS	F	P
isl	1	0.0002836	0.0002836	0.09	0.798
sex	1	0.0479273	0.0479273	10.18	0.086
year	2	0.0024518	0.0012259	2.21	0.116
isl*sex	1	0.0017768	0.0017768	2.43	0.259
isl*year	2	0.0066462	0.0033231	5.99	**0.004**
sex*year	2	0.0094182	0.0047091	8.48	**0.000**
isl*sex*year	2	0.0014634	0.0007317	1.32	0.273
Error	84	0.0466344	0.0005552		
Total	95	0.1166017			

These results are the same as for the unrestricted model for all effects not including the year effect. For effects including year, the results are quite different; in particular, the interactions of year with sex and island are highly significant.

Analysis of Effects

In this case it is not obvious what further analyses would make sense and be interesting. The male and female means could be compared, as in the unrestricted model above, but the low significance of this effect and the highly significant sex x year interaction make the meaningfulness of this comparison questionable. It might be more appropriate to estimate the variance components for the sex x year and island x year interactions; these estimates are in the printout of EMSes above. It is not clear what these estimates tell us, except that these sources of variation, while quite significantly non-zero, appear less important than the within-sample variability.

Unbalanced Data Set

Preliminary Description of the Data

Summary Statistics

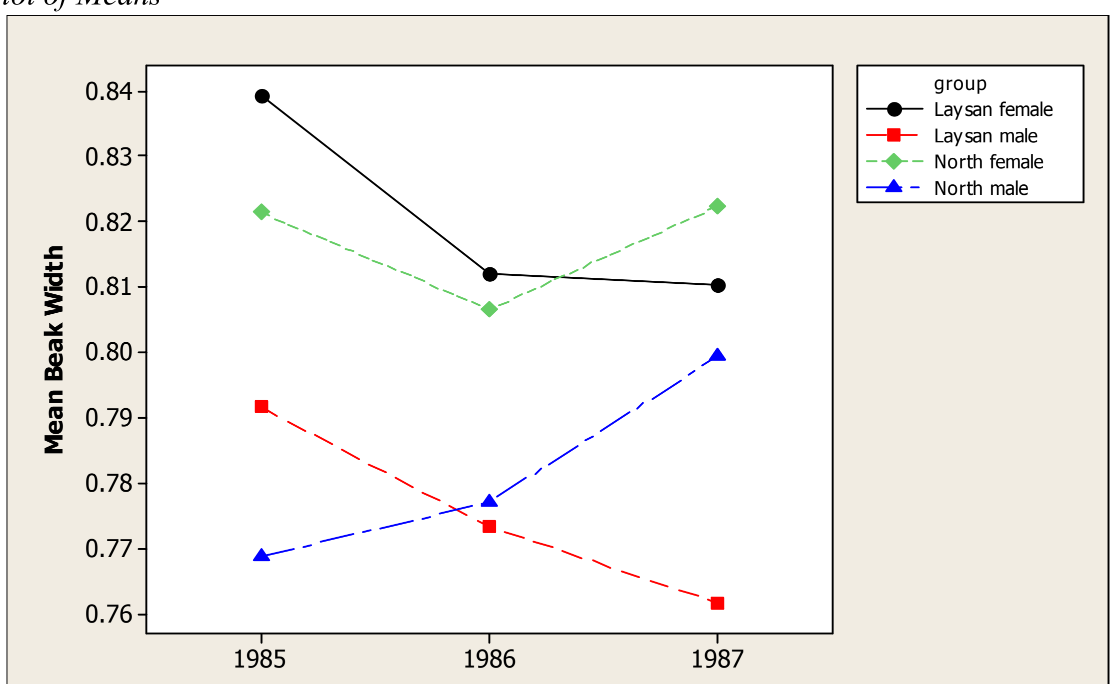
		males			females	
	1985	1986	1987	1985	1986	1987
Laysan	$\hat{Y}$: 0.839 *n*: 125	0.812 102	0.810 135	0.792 89	0.773 71	0.762 69
North	0.821 21	0.807 35	0.822 14	0.769 8	0.777 15	0.799 10

There are many more observations from Laysan than North Island, so the former would dominate any unweighted comparisons combining islands (*e.g.* simple means for sexes or years). In the absence of a difference between islands this might have little effect.

There also are more observations for males than for females, and this imbalance is greater on North Island than on Laysan. Given the substantial beak width difference between the sexes, this imbalance could have a major effect on comparisons among years or especially islands (*e.g.* the raw mean for North Island is increased, relative to that for Laysan, by the greater proportion of males).

The sample means show the same general pattern as in the balanced subset of the data analyzed earlier: male beaks are consistently wider than female beaks, but differences between islands vary among years and between sexes.

Plot of Means



The pattern of the means can be seen better in the scatterplot on the previous page. Because the balanced subsample used previously included most of the observations from North island, the North island means are not greatly different in the full data set. The Laysan means, though, are somewhat different than those for the balanced subset; in particular the trends for males and females over the three years are fairly similar (the lines a roughly parallel). That the trend between 1985 and 1986 differed between males and females on North island, but not on Laysan, suggests a three-way interaction.

ANOVA for Fixed-Effects Model:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
island	1	0.000000	0.000119	0.000119	0.21	0.648
sex	1	0.291335	0.114475	0.114475	200.59	**0.000**
year	2	0.085347	0.008877	0.004438	7.78	**0.000**
island*sex	1	0.003896	0.001775	0.001775	3.11	0.078
island*year	2	0.018536	0.021542	0.010771	18.87	**0.000**
sex*year	2	0.003117	0.003615	0.001807	3.17	**0.043**
island*sex*year	2	0.002455	0.002455	0.001227	2.15	0.117
Error	682	0.389206	0.389206	0.000571		
Total	693	0.793893				

Because the data are unbalanced, we will use the adjusted (=Type III) SSes. As the printout above shows, these SSes are quite different than the sequential SSes for some effects, most notably for the year effect, which is an order of magnitude smaller after being “adjusted” for the various interactions.

The conclusions for the full data set are basically consistent with those for the balanced subset analyzed above. Specifically, the sex effect and island x year interactions again are highly significant, the island main effect is quite nonsignificant, and the island x sex interaction is of uncertain significance. The greatest difference is that in the full data set the year effect is highly significant. Interestingly, the sex x year interaction, which is so clear in the plots of the balanced subsample, is considerably less significant in the full data set.

ANOVA for (Unrestricted) Mixed Model

EMSes

```
ISLAND             Var(Error) + 23.795 Var(ISLAND*SEX*YEAR)
                   + 47.59 Var(ISLAND*YEAR) + Q(ISLAND,ISLAND*SEX)

SEX                Var(Error) + 23.795 Var(ISLAND*SEX*YEAR)
                   + 47.59 Var(SEX*YEAR) + Q(SEX,ISLAND*SEX)

ISLAND*SEX         Var(Error) + 23.795 Var(ISLAND*SEX*YEAR)
                   + Q(ISLAND*SEX)

YEAR               Var(Error) + 24.311 Var(ISLAND*SEX*YEAR)
                   + 48.622 Var(SEX*YEAR) + 48.622 Var(ISLAND*YEAR)
                   + 97.244 Var(YEAR)

ISLAND*YEAR        Var(Error) + 24.311 Var(ISLAND*SEX*YEAR)
                   + 48.622 Var(ISLAND*YEAR)

SEX*YEAR           Var(Error) + 24.311 Var(ISLAND*SEX*YEAR)
                   + 48.622 Var(SEX*YEAR)

ISLAND*SEX*YEAR    Var(Error) + 24.311 Var(ISLAND*SEX*YEAR)
```

The important points to notice about these EMSes are that: (1) the constants weighting the variance components are not integers, and may differ among EMSes for a given variance component; and (2) the quadratic forms for fixed main effects include interaction effects as well as the main effect. These two features make tests more complicated than in the balanced case.

The first complication is that *quasi-F* tests are needed in cases which ordinarily would not require them. For instance, in the balanced case the three-way interaction MS is the appropriate error term to test the island x sex interaction. But now, because of differences between the two MSes in the constants for the island x sex x year variance component, a composite error term is required:

```
Source: ISLAND*SEX
Error: 0.9788*MS(ISLAND*SEX*YEAR) + 0.0212*MS(Error)

                      Denominator   Denominator
DF     Type III MS             DF            MS       F Value    Pr > F
 1    0.0017745315           2.04    0.00121336         1.462    0.3480
```

The second complication is that main effects of the fixed factors cannot be tested without assuming there is no island x sex interaction:

```
Source: ISLAND *
Error: 0.9788*MS(ISLAND*YEAR) + 0.0212*MS(Error)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 1    0.0001188694          2.00  0.0105544302        0.011    0.9251
* - This test assumes one or more other fixed effects are zero.

Source: SEX *
Error: 0.9788*MS(SEX*YEAR) + 0.0212*MS(Error)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 1    0.1144746904          2.03  0.0017812377       64.267    0.0146
* - This test assumes one or more other fixed effects are zero.
```

Effects including the random factor also require *quasi-F* tests, but these have the same integer weightings as in the balanced case:

```
Source: YEAR
Error: MS(ISLAND*YEAR) + MS(SEX*YEAR) - MS(ISLAND*SEX*YEAR)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 2    0.0044383374          2.13  0.0113511511        0.391    0.7167


Source: ISLAND*YEAR
Error: MS(ISLAND*SEX*YEAR)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 2    0.0107709573             2  0.0012272984        8.776    0.1023

Source: SEX*YEAR
Error: MS(ISLAND*SEX*YEAR)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 2    0.0018074921             2  0.0012272984        1.473    0.4044


Source: ISLAND*SEX*YEAR
Error: MS(Error)
                      Denominator   Denominator
DF     Type III MS            DF             MS      F Value    Pr > F
 2    0.0012272984           682  0.0005706836        2.151    0.1172
```

The conclusions for this mixed model are essentially the same as for the mixed-model analysis of the balanced subset of the data: the only significant effect is the difference between the sexes (which in this case definitely is significant, in contrast to the borderline result for the smaller data set).