Model Based Statistics in Biology.

Part III. The General Linear Model.

Chapter 10.3 One way ANOVA, Fixed Effects

ReCap.	Part I (Chapters 1,2,3,4)
ReCap	Part II (Ch 5, 6, 7)
ReCap	Part III (Ch 9)
10.1	Single Sample t-test
10.2	Two Sample t-test
10.3	One way ANOVA, Fixed Effects
10.4	One way ANOVA, Random Effects

on chalk board

"Let the computer find out" is a poor strategy for researchers who do not bother to think clearly about the problem of interest and its scientific setting. The sterile analysis of 'just the numbers' will continue to be a poor strategy for progress in the sciences.

p117 in Burnham, K.P., D.R. Anderson. 1998. Model Selection and Inference. NewYork: Springer.

ReCap Part I (Chapters 1,2,3,4)
Quantitative reasoning: Example of scallops,
which combined models (what is the relation of scallop density to substrate?)
with statistics (how certain can we be?)

ReCap Part II (Chapters 5,6,7)
Hypothesis testing uses the logic of the null hypothesis to make a decision about an unknown population parameter.
Estimation is concerned with the specific value of an unknown population parameter.

ReCap (Ch 9) The General Linear Model is more useful and flexible than a collection of special cases.
Regression is a special case of the GLM. We have seen an examples with the explanatory variable X fixed, with the explanatory measured with error, and for a non-linear (exponential and power law) relations of response to explanatory variable. **ReCap** (Ch 10) ANOVA a special case of the general linear model.
Explanatory variable is on nominal scale.
Special case of one-way (single factor) ANOVA: two means. Called a t-test.

Today: ANOVA as a special case of the GLM.
Single Factor ANOVA - Fixed Effects

Wrap-up. GLM. ANOVA. Explanatory variable on nominal scale.

Relation of response to explanatory variable expressed as set of means, rather than as a slope. Factor fixed by experimental design. We carry out planned comparisons of classes, based on our knowledge of the reasons for collecting the data.

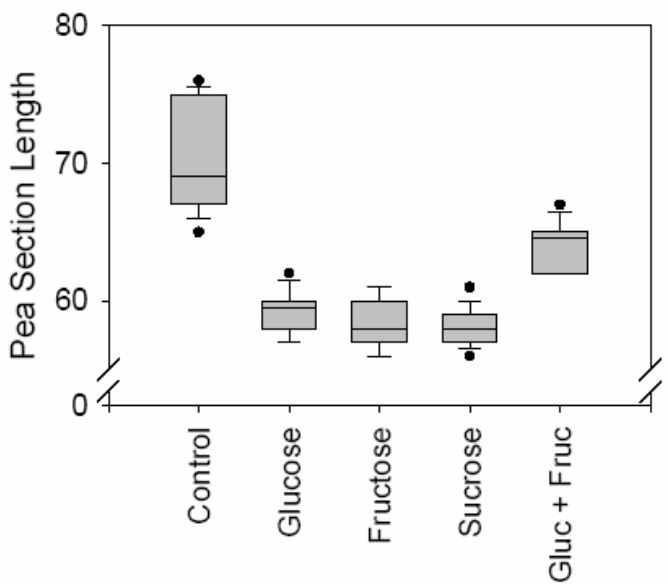
Application of GLM. One way ANOVA, fixed effects.

Pea section growth data, from Box 9.4 in Sokal and Rohlf (1995).
Does growth depend on treatment (control versus 4 different sugars with auxin present) ?

1. Construct model

Verbal model. Pea section length in treated groups differ from the control (untreated) group.

Graphical model —>



The response variable (*Len* = pea section length) is a measure of growth, expressed in ocular units (0.114 ocular units per mm).

Explanatory variable is treatment *Trt*. (Control and 4 different media).
The explanatory variable is on a nominal type of scale (classes).
The explanatory variable is called a fixed effect. We are interested in these particular sugars, and whether than change growth relative to the untreated control. We do not view the sugars as random samples of a variety of sugars, either singly or in combination. Later, we will see examples of random effect explanatory variables.

Formal model (write GLM)

$$Len = \beta_o + \beta_{Trt} \cdot Trt + \varepsilon$$

2. Execute analysis.

Place data in model format:

Column with response variable, *Len* = pea section length
Column with explanatory variable, *Trt* = 0 (control), 1, 2, 3, 4
These are labels (categories), not numbers on ratio scale.

2. Execute analysis.

Code model statement in statistical package according to the GLM

$$Len = \alpha + \beta_{Trt} \cdot Trt + \varepsilon$$

```
MTB> ANOVA 'Len' = 'Trt'
MTB> GLM 'Len' = 'Trt'
SUBC> fits c4;
SUBC> res c5.
```

Fits and residuals from:

a. model statement output of fitted values and residuals (as above)

b. direct calculation of parameters (five means)

c. parameters reported by GLM routine

β_o = 61.94 ocular units = 7.06 mm

$\beta_o + \beta_{Trt}$ = [70.1 59.3 58.2 58.0 64.1] ocular units

3. Evaluate model.

Evaluate straight line assumption.

No straight lines (regression) used, so skip this.

Evaluate error model. Plot residuals vs fitted values.

```
MTB > plot 'res' 'fits'
res      -
         -                                                    *
         -                                                    2
       4.0+
         -      *2                    *
         -           *
         -       *   *                *
         -      2    3                3                      *
       0.0+     33   2                *                      *
         -      32                    *
         -           *
         -      *2   2                3                      *
         -                                                   3
      -4.0+
         -                                                   *
         -
         -
           --+---------+---------+---------+---------+---------+----fits
          57.5       60.0       62.5       65.0       67.5       70.0
```

Homogeneity?

Nearly so. Residuals extend over slightly wider range in control than treated groups (see figure), but not substantially so.

Normal errors ?

Histogram looks normal.

Confirmed by nscores (not shown), nearly a straight line.

```
MTB > hist c9;
 SUBC> incr 2.
 Histogram of res   N = 50

 Midpoint   Count
    -6.00       1  *
    -4.00       3  ***
    -2.00      13  *************
     0.00      20  ********************
     2.00       9  *********
     4.00       3  ***
     6.00       1  *
```

3. Evaluate error model.

It is interesting to note that if we attempt to check the assumptions on 'the data' before computing the residuals, the assumptions appear to be violated. In particular the distribution of the data deviates markedly from normal. In this case, evaluating the response variable leads to the wrong conclusion.

```
MTB > hist c6
Histogram of len   N = 50

Midpoint   Count
      56       3  ***
      58      14  **************
      60       8  ********
      62       8  ********
      64       2  **
      66       5  *****
      68       5  *****
      70       1  *
      72       1  *
      74       0
      76       3  ****
```

Independent errors? Plot of residuals vs neighbors (lag 1) hints at some degree of dependence (not shown). But not strong.

In this example we were looking for large violations ($30 < n < 100$). There were none. As we will see with the analysis of the tick scutum width data, checking the assumptions before computing the residuals would lead us to try to correct a problem that does not exist. A medical analogy would be chemotherapy for a lump, with determining is the lump is benign or cancerous.

4. State population and whether sample is representative.

Because this is a fixed effect model, we are inferring to the population differences for these four treatments only. We are not inferring to other sugars or combinations of sugars.

Population is all possible measurements, given the method of applying treatments and the protocol for taking measurements. It is taken to be representative (not biased).

5. Decide on mode of inference. Is hypothesis testing appropriate?

In ANOVA with fixed effects in multiple groups we start with a test of the overall hypothesis that there is variance among groups. We start with this because it uses all the data and hence is the most sensitive (able to detect differences). If there are differences, then we examine the parameters of biological interest - in this case whether the treated groups differ from the untreated groups.

6. H_A / H_o pair for overall test.

There is one term in the model. Is this term significant ? (not due to chance).
The research hypothesis is that the treatments differ in effect.

$$H_A: \mu_{Control} \neq \mu_{Glucose} \neq \mu_{Fructose} \neq \mu_{Sucrose} \neq \mu_{Glucose+Fuctose}$$

The null hypothesis is the treatments do not differ in effect.

$$H_o: \mu_{Control} = \mu_{Glucose} = \mu_{Fructose} = \mu_{Sucrose} = \mu_{Glucose+Fuctose}$$

This is equivalent to the following pair of hypotheses.

The symbol $\beta_{Trt} \cdot Trt$ stands for the collection of 5 means.
If the means differ, then var($\beta_{Trt} \cdot Trt$) >0 H_A: var($\beta_{Trt} \cdot Trt$) > 0
If the means are the same,
then there is no variance in $\beta_{Trt} \cdot Trt$ H_o: var($\beta_{Trt} \cdot Trt$) = 0

H_A/H_o are statements about parameters of the population. They are not statements about the estimates of the parameters, which we will calculate from the data.

When we have a fixed effects model, it is natural to investigate the sources of differences among groups, if the overall test reveals a difference.
For experiments with fixed effects, we usually have some expectations about the direction and sometimes the strength of the contrasts among groups. Based on these, we can undertake planned or *a priori* (before the fact) comparisons. Such comparisons use our biological understanding more effectively than unplanned or *a posteriori* (after the fact) comparisons. We can set up planned comparisons any way we like, but we need to take into account the number of comparisons.

6. H_A / H_o pairs for planned comparisons.

In this example the most natural comparison is the control versus the other four treatments (one degree of freedom used).
A second comparison is mixed sugars versus pure sugars (one degree of freedom used).
A third comparison is within the pure sugars: polysaccharides (sucrose) versus monosaccharides (glucose and fructose).

Here are the three planned comparisons, showing the research hypothesis H_A only.
$H_A: \beta_{control} \neq (1/4)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose} + \beta_{gluc+fruc})$ [control versus treatment]
$H_A: (\beta_{gluc+fruc}) \neq (1/3)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose})$ [mixed versus pure]
$H_A: (\beta_{sucrose}) \neq (1/2)(\beta_{glucose} + \beta_{fructose})$ [polysaccharides vs monosaccharides]

This is not the only way to undertake planned comparisons. Other sequences are possible.

6. H_A / H_o pair, test statistic, distribution, tolerance of Type I error.

State test statistic F-ratio

GLM	$Len - \beta_o$	=	$\beta_{Trt} \cdot Trt$	+	ε
Source	Total	=	Trt	+	Resid

Distribution of test statistic F-distribution

Tolerance for Type I error 5% (conventional level

Tolerance for Type I error is again set at the conventional 5% level but to achieve this, we need to take into account the number of comparisons. For example, if we undertake 20 comparisons, we expect on average one false acceptance of H_o (5% = one in 20). If we use a fixed criterion of 5% for multiple tests, then we expect the error rate after multiple tests (called the ***experimentwise error rate***) to exceed the 5% level. This suggests that we should lower the criterion level according to the number of tests, so that the experimentwise error rate does not rise above our stated tolerance for Type I error. One solution is to divide the tolerance level by the number of tests. Thus with 3 tests we would use 5% / 3 = 1.67% to declare significance. This is called the Bonferroni method for limiting the experimentwise error rate. An alternative method is called the Dunn-Sidak method (Ury 1976 in Sokal and Rohlf 1995). For k tests, the experimentwise error rate is

$\alpha_{\exp wise} = 1 - (1 - \alpha)^k$ hence we use $\alpha = 1 - (1 - \alpha_{\exp wise})^{1/k}$ for our error rate.

For three tests we would use $\alpha = 1 - (1 - 0.05)^{1/3} = 0.017$ to achieve a 5% error rate.

7. ANOVA - Calculate df and partition according to model.

Compute total degrees of freedom $df_{total} = n - 1 = 50 - 1 = 49$

Partition df_{total} according to model, using rules

5 levels in factor called treatment *Trt* $df_H = 5 - 1 = 4$

$df_{res} = df_{total} - df_{Trt}$ $df_{res} = 49 - 4 = 45$

Calculate variance and partition according to model.

Compute $SS_{tot} = Var(Len) \cdot df_{total}$

$SS_{tot} = 26.996 \cdot 49 = 1322.8$

Partition SS_{tot} (model format)

	$Len - \beta_o$	=	$\beta_{Trt} \cdot Trt$	+	ε
Source	Total	=	Trt	+	resid
	49	=	4	+	45
MTB> ANOVA	'Len'	=	'Trt'		
MTB> GLM	'Len'	=	'Trt';		
SS	1322.82	=	SS_{Trt} + SS_{res}		

7. ANOVA

Move Source, df, and SS to ANOVA table. Compute MS and F to complete table.

Source	DF	SS	MS	F	P
trt	4	1077.3	269.330	49.37	0.000
Error	45	245.5	5.456		
Total	49	1322.8	[26.996]		

Calculate Type I error from F-distribution.

$p < 0.0001$

```
MTB > cdf 49.33 k1;
SUBC> F 4 45.
     49.33    1.00000
```

8. Recompute p-value if necessary.

Assumptions met, skip this step

9. Declare decision about model terms.

$p < 0.001$

$p < \alpha$ so reject H_o and accept H_A

Accept H_A: $\mu_{Control} \neq \mu_{Glucose} \neq \mu_{Fructose} \neq \mu_{Sucrose} \neq \mu_{Glucose+Fuctose}$

Reject H_o: $\mu_{Control} = \mu_{Glucose} = \mu_{Fructose} = \mu_{Sucrose} = \mu_{Glucose+Fuctose}$

Pea section length differs significantly among the 5 groups (control and 4 treatments). $F_{4,45} = 49.37$, $p < 0.0001$

10. Report and interpret parameters of biological interest

Where are the differences, among the 5 groups ?

Two approaches *A priori* Planned comparisons. Use prior knowledge.

A posteriori. Hunting. See Burnham and Anderson quote.

Statistical packages tend to foster the use of *a posteriori* comparisons.

In most situations, one can develop *a priori* comparisons that use information more effectively. Often these are one-tailed (directional) comparisons and hence have more capacity (power) to detect differences. It is best to undertake planned comparisons. If this fails, only then switch to exploratory approach based on *a posteriori* analyses.

We look at the three *a priori* comparisons. To achieve an experimentwise error rate of 5% will use $\alpha = 0.017$ from the Dunn-Sidak formula.

10. Report and interpret parameters of biological interest

First planned comparison: Growth in treated media differs from that in untreated.

H_A: $\beta_{control} \neq (1/4)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose} + \beta_{gluc+fruc})$

When we compare control versus the 4 treatments, the resulting ANOVA table uses the mean square error based on two groups, rather than on 5 groups (as in previous analysis).

Source	DF	SS	MS	F
df1	1	832.3	832.32	81.45
Error	48	490.5	10.22	
Total	49	1322.8	27.00	

```
MTB > let k1 = 245.5/45
MTB > let k2 = 832.32/k1
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k2 k4
K2        152.564   # F-ratio
K4        0         # p-value
```

For *a priori* comparisons we compute the F-ratio based on the MS estimate from 5 groups with 45 degrees of freedom.

The strength of the difference is of interest so we compute the 95% confidence limits and compare the means.

sugar suppresses growth by (70.1 – 59.9)/70.1 = 15%

```
MTB > invcdf .975 k5;
SUBC> t 45.
MTB > print k5
K5        2.01409
MTB > let k6 = 59.9 - k5*.472
MTB > let k7 = 59.9 + k5*.472
MTB > let k8 = 70.1 - k5*1.26
MTB > let k9 = 70.1 + k5*1.26
MTB > print k6-k9
K6        58.9494     # lower 95% CI
K7        60.8507     # upper 95% CI
K8        67.5622     # lower 95% CI
K9        72.6377     # upper 95% CI
```

10. Report and interpret parameters of biological interest

Second planned comparison: Growth in mixed versus differs from that in pure sugar.

H_A: $(\beta_{gluc+fruc}) \neq (1/3)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose})$

Source	DF	SS	MS	F
df2	1	48.13	48.133	6.11
Error	38	299.47	7.881	
Total	39	347.60	8.913	

We compare the two groups (one with 10 and one with 30 observations) to obtain the MS due to the two groups.

```
MTB > let k2 = 48.133/k1
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k2 k4
K2        8.82275
K4        0.00475895
```

This MS is used to recompute the F-ratio relative to the MS error from all 5 groups.
The F-ratio is 8.82, not 6.11, as in the ANOVA table.

The confidence limits are on the order of +/– 2 units.

```
MTB > print k6 - k9
K6   59.3870      # lower 95% CI  pure sugar
K7   61.6790      # upper 95% CI
K8   57.0997      # lower 95% CI  mixed sugar
K9   58.9003      # upper 95% CI
```

Mixed glucose + fructose reduces growth by (60.533 – 58)/60.533) = 4% relative to pure sugars

Third planned comparison: Growth in polysaccharide differs from that in monosaccharides.

H_A: $(\beta_{sucrose}) \neq (1/2)(\beta_{glucose} + \beta_{fructose})$

Source	DF	SS	MS	F	P
df3	1	190.82	190.817	58.94	0.000
Error	28	90.65	3.237		
Total	29	281.47	9.706		

```
MTB > let k1 = 245.5/45
MTB > let k2 = 190.817/k1
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k3 k4
k3        34.9766
K4        0.000000417
```

The monosaccharides suppress growth more than the polysaccharide (sucrose).

```
MTB > describe c11;
 SUBC> by c12.
      df3         N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
 Gluc       1   10   59.300   59.500   59.250    1.636    0.517
 Fruc       2   10   58.200   58.000   58.125    1.874    0.593
 Sucr       3   10   64.100   64.500   64.000    1.792    0.567

 MTB > aoveoneway c22 c23 c25
                                    INDIVIDUAL 95 PCT CI'S FOR MEAN
                                    BASED ON POOLED STDEV
 LEVEL        N      MEAN     STDEV  --+---------+---------+---------+----
 glucose     10    59.300     1.636       (---*----)
 fructose    10    58.200     1.874  (----*---)
 sucrose     10    64.100     1.792                       (---*----)
                                     --+---------+---------+---------+----
 POOLED STDEV =    1.770           57.5      60.0      62.5      65.0
```

10. Report and interpret parameters of biological interest

Conclusions from the 3 planned comparisons of parameters (means).

A 2% sugar solution reduces growth.

$F_{1,45} = 152.564 \quad \alpha = 0.017, p < 0.0001$

Mixed glucose + fructose reduces growth relative to pure sugars

$F_{1,45} = 8.82 \quad \alpha = 0.017 > p = 0.00476$

The monosaccharides (fructose, glucose) suppress growth more than the polysaccharide (sucrose).

$F = 34.98 \quad \alpha = 0.017 > p = 0.00000417$

Conclusions based on confidence limits for the 3 planned comparisons.

A 2% sugar solution (4 groups) reduces growth by (70.1 – 59.9)/70.1 = 15%

Control: mean(*Len*) = 70.1 units, 95% CI = 67.6 to 72.6units, n = 10

With sugar: mean(*Len*) = 59.9 units, 95% CI = 58.9 to 60.9 units, n = 40

Mixed glucose + fructose reduces growth by (60.533 – 58)/60.533) = 4% relative to pure sugars.

Mixed: mean(*Len*) = 58 units, 95% CI = 57.1 to 58.9, units, n = 10

Pure sugar: mean(*Len*) = 60.5 units, 95% CI = 59.4 to 61.8 units, n = 30

Sucrose (a polysaccharide) reduces growth less than glucose or fructose (monosaccharides).

Sucrose: mean(*Len*) = 64.1 units, 95% CI = 62.6 to 65.6 units, n = 10

Fructose: mean(*Len*) = 58.2 units, 95% CI = 56.7 to 59.7 units, n = 10

Glucose: mean(*Len*) = 59.3 units, 95% CI = 57.8 to 60.8 units, n = 10

For these three planned comparisons, confidence limits do not overlap and so conclusions were readily drawn from examination of the confidence limits.