Model Based Statistics in Biology.
Part IV. The General Linear Model. Multiple Explanatory Variables.
Chapter 13.1 Fixed Effects ANOVA (no interactive effects)

ReCap.	Part I (Chapters 1,2,3,4), Part II (Ch 5, 6, 7)
ReCap	Part III (Ch 9, 10, 11)
ReCap	Multiple Regression (Ch 12)
13.1	Fixed Effects ANOVA (no interactive effects)
13.2	Fixed Effects ANOVA (interactive effects)
13.3	Fixed*Random Effects (Paired t-test)
13.4	Fixed*Random Effects (Randomized Block)
13.5	Fixed*Random Effects (Repeated Measures)
13.6	Nested Random Effects (Hierarchical ANOVA)
13.7	Random within Fixed (Hierarchical ANOVA)
13.8	More Than Two Factors (to be written)

Ch13.xls
Limpet Respiration

on chalk board

ReCap Part I (Chapters 1,2,3,4) Quantitative reasoning is based on models, including statistical analysis based on models.
ReCap Part II (Chapters 5,6,7)
Hypothesis testing uses the logic of the null hypothesis to declare a decision.
Estimation is concerned with the specific value of an unknown population parameter.
ReCap (Ch 9, 10,11) The General Linear Model with a single explanatory variable.
ReCap (Ch 12) GLM with more than one regression variable (multiple regression)

Today: Two-way ANOVA.
One response variable Y as a function of two explanatory variables X_1 X_2. Both explanatory variables are categorical, on a nominal scale.

Wrap-up. General Linear Model with two classification variables,
i.e. two explanatory variables on a nominal scale.

New concept, the interaction term.

Example. GLM, applied to 2-way ANOVA in Sokal and Rohlf 1995 p332.
Does oxygen consumption depend on salinity, in two species of limpet ?

1. Construct model

Data are: oxygen consumption (microliters per minute) / (mg dry weight) of two species of limpet, at three different salinities.

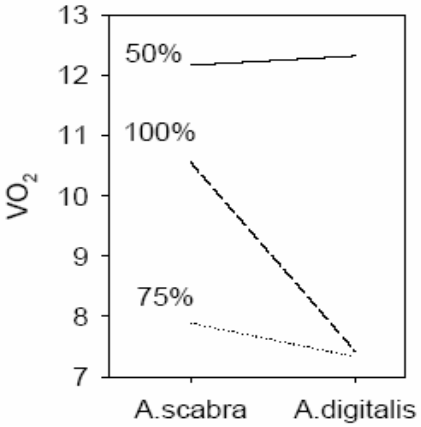
Response variable
Oxygen consumption. $_{O2}$ = µl O_2 min^{-1} mg^{-1} dry weight (ratio scale)

Explanatory variables are salinity levels and species.
Species. X_{sp} = *A.scabra*, *A.digitalis* (two categories, nominal scale)

Salinity. X_{sal} = salinity in three categories: 100%, 75%, 50%
This quantity was measured on a ratio scale, but it is here reduced to nominal scale, of three categories.

Verbal model. Oxygen consumption depends on salinity and species

Graphical model.
Y-axis = O_2 consumption, from 0 to 14
X-axis = 50% 75% 100%
Connect 3 means of each of two species.
A. scabra 12.17 7.89 10.56
A. digitalis 12.33 7.35 7.43



Graph suggests that effect of salinity on O_2 consumption may depend on species of limpet.

Formal Model

Write GLM: $V = \beta_o + \beta_{sp} \cdot X_{sp} + \beta_{sal} \cdot X_{sal} + \beta_{sp\ x\ sal} \cdot X_{sp} \cdot X_{sal} +$ residual
S&R95 $V_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \cdot \beta)_{ij} + \epsilon_{ijk}$

A model has been written using two forms of notation. One is typical notation for the GLM, the other is from Sokal and Rohlf 1995. The symbols in each of the two sets of notation have been aligned, to facilitate translation.

The two formal models have the same form, despite the differences in notation. This formal structure is required in order to use a computer to carry out the analysis.

There are three explanatory terms, one for salinity, one for species, and one for interactive effects--the dependence of salinity effects on species. Graphical interpretation is that the shape of the relation of O_2 consumption to salinity in one species does not match the shape of the same relation in the other species. The sample data suggests different patterns of response to salinity in the two species.

```
MTB> ANOVA 'VO2' = 'Xsp'  'Xsal'  'Xsp'*'Xsal'
MTB> GLM   'VO2' = 'Xsp'  'Xsal'  'Xsp'*'Xsal'
SUBC> fits c4;
SUBC> res c5.
```

2. Execute analysis.

Place data in model format:

Column labelled VO2, with response variable oxygen consumption
Column labelled X_{sal} with explanatory variable, X_{sal} = 0 or 1
Column labelled X_{sp} with explanatory variable, X_{sp} = 0 or 1
These are labels (categories), not numbers on ratio scale.

Code model statement in statistical package according to the GLM

$$= \beta_o + \beta_{sp} \cdot X_{sp} + \beta_{sal} \cdot X_{sal} + \beta_{sp\ x\ sal} \cdot X_{sp} \cdot X_{sal} + \epsilon$$

The grand mean. $_0 = 48^{-1} \Sigma = 48^{-1} \cdot 461.74 = 9.62\ \mu l\ min^{-1}\ mg^{-1}$

The fitted values are those for each of the 6 cells.

$Mean(_{sal=50,\ sp=As}) = 8^{-1} \Sigma(_{sal=50,\ sp=As}) = 8^{-1} \cdot 97.39 = 12.17\ \mu l\ min^{-1}\ mg^{-1}$

	A.scabra	A.digitalis	Sums		
50%	97.39	98.61	196.00		
75%	63.12	58.70	121.82		
100%	84.49	59.43	143.92		
Sums	245.00	216.74	461.74		

	A.scabra	A.digitalis	Means		
50%	12.17	12.33	12.25		
75%	7.89	7.34	7.61		
100%	10.56	7.43	9.00		
Means	10.21	9.03	9.62	=βo	

	A.scabra	A.digitalis	Means-βo	βsal	
50%	2.55	2.71	2.63		
75%	-1.73	-2.28	-2.01	-2.3181	=(Mean@100%-βo)+(Mean@75%-βo)
100%	0.94	-2.19	-0.62	-0.3123	=(Mean@100%-βo)/2
Means-βo	0.59	-0.59			
		βsp			

	A.scabra	A.digitalis	Means-Σβ	βsal*Sp	
50%	-0.665	0.665	0.00		
75%	-0.313	0.313		-0.1762	=(Mean@100%-Σβ)+(Mean@75%-Σβ)
100%	0.977	-0.977		-0.4887	=(Mean@100%-Σβ)/2
Means-Σβ	0.00	0.00			

Next, we compute the means for each level of each factor.

$Mean(_{sal=50}) = 16^{-1} \Sigma_{sal=50} = 16^{-1} \cdot 196 = 12.25\ \mu l\ min^{-1}\ mg^{-1}$
$Mean(_{sal=75}) = 16^{-1} \Sigma_{sal=75} = 16^{-1} \cdot 121.82 = 7.61\ \mu l\ min^{-1}\ mg^{-1}$
$Mean(_{sal=100}) = 16^{-1} \Sigma_{sal=100} = 16^{-1} \cdot 143.92 = 9.00\ \mu l\ min^{-1}\ mg^{-1}$
$Mean(_{sp=As}) = 24^{-1} \Sigma_{sp=As} = 24^{-1} \cdot 245 = 10.21\ \mu l\ min^{-1}\ mg^{-1}$
$Mean(_{sp=Ad}) = 24^{-1} \Sigma_{sp=Ad} = 24^{-1} \cdot 216.74 = 9.03\ \mu l\ min^{-1}\ mg^{-1}$

2. Execute analysis.

Then we adjust these means for the grand mean.

$\text{Mean}_{(sal=50)} - \hat{\beta}_0 = 12.25 - 9.62 = 2.63\ \mu l\ min^{-1}\ mg^{-1}$

$\text{Mean}_{(sal=75)} - \hat{\beta}_0 = 7.61 - 9.62 = -2.01\ \mu l\ min^{-1}\ mg^{-1}$

$\text{Mean}_{(sal=100)} - \hat{\beta}_0 = 9.00 - 9.62 = -0.62\ \mu l\ min^{-1}\ mg^{-1}$

$\text{Mean}_{(sp=As)} - \hat{\beta}_0 = 10.21 - 9.62 = 0.589\ \mu l\ min^{-1}\ mg^{-1}$

$\text{Mean}_{(sp=Ad)} - \hat{\beta}_0 = 9.03 - 9.62 = -0.589\ \mu l\ min^{-1}\ mg^{-1}$

$\text{AdjMean}_{(sal=75,\ Sp=Ad)} = \text{Mean}_{(sal=75,\ Sp=Ad)} - \hat{\beta}_0 - \text{Mean}_{(sal=75)} - \text{Mean}_{(sp=Ad)}$

$= 7.34 - 9.62 - (7.61 - 9.62) - (9.03 - 9.62) = 3.13\ \mu l\ min^{-1}\ mg^{-1}$

$\text{AdjMean}_{(sal=100,\ Sp=Ad)} = \text{Mean}_{(sal=100,\ Sp=Ad)} - \hat{\beta}_0 - \text{Mean}_{(sal=100)} - \text{Mean}_{(sp=Ad)}$

$= 7.43 - 9.62 - (9.00 - 9.62) - (9.03 - 9.62) = -0.977\ \mu l\ min^{-1}\ mg^{-1}$

The GLM parameters are linear combinations of adjusted means.

$\hat{\beta}_{sal=100} = (\text{Mean}_{(sal=75)} - \hat{\beta}_0) + (\text{Mean}_{(sal=100)} - \hat{\beta}_0)/2$

$= -2.01 - 0.62 / 2 = -2.31\ \mu l\ min^{-1}\ mg^{-1}$

$\hat{\beta}_{sal=75} = (\text{Mean}_{(sal=100)} - \hat{\beta}_0)/2 = -3.26\ \mu l\ min^{-1}\ mg^{-1}$

$\hat{\beta}_{Sp=Ad} = \text{Mean}_{(sp=Ad)} - \hat{\beta}_0 = -0.589\ \mu l\ min^{-1}\ mg^{-1}$

$\hat{\beta}_{Ad*100} = \text{AdjMean}_{(sal=100,Sp=Ad)}/2 = -0.977/2 = -0.489\ \mu l\ min^{-1}\ mg^{-1}$

$\hat{\beta}_{Ad*75} = \text{AdjMean}_{(sal=75,\ Sp=Ad)} + \text{AdjMean}_{(sal=100,Sp=Ad)}/2$

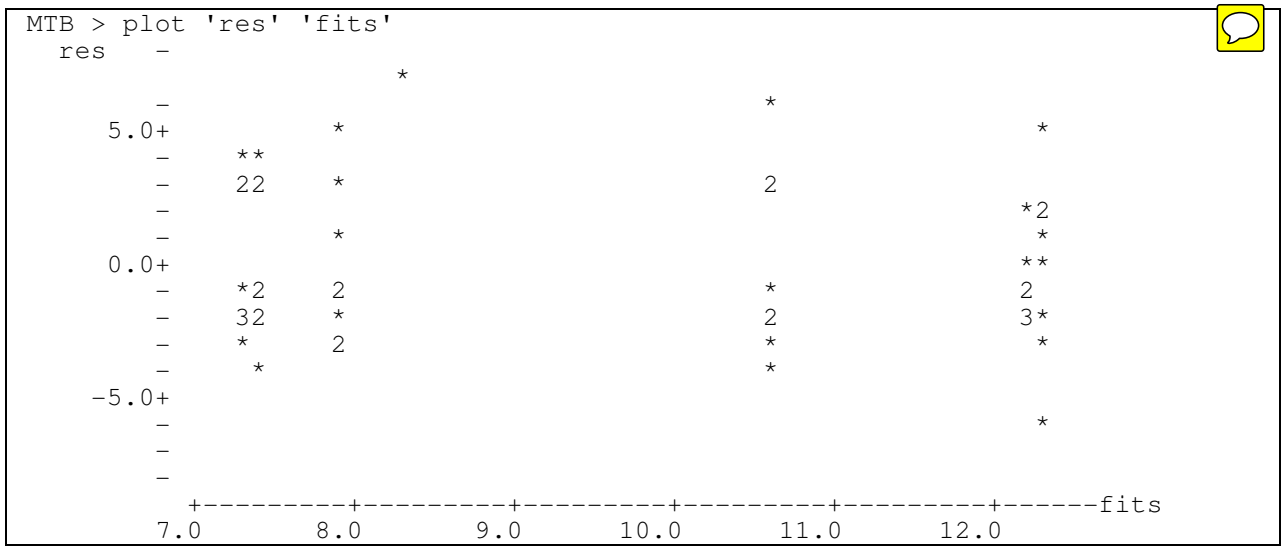
$= 3.13 - 0.489 = -0.176\ \mu l\ min^{-1}\ mg^{-1}$

Here are GLM parameters produced by a statistical package (SPlus)

	Value	Std. Error	t value
(Intercept)	9.6195833	0.4462817	21.5549605
SP	-0.5887500	0.4462817	-1.3192342
Sal1	-2.3181250	0.5465812	-4.2411359
Sal2	-0.3122917	0.3155688	-0.9896152
SPSal1	-0.1762500	0.5465812	-0.3224590
SPSal2	-0.4887500	0.3155688	-1.5487907

3. Evaluate the model.

Plot residuals versus fits (cell means).



a. No line fitted in model, so skip this evaluation.

b. No systematic change in residuals with increase in fitted values (*i.e.* no cones) so residual homogeneous, no need to revise error structure of model.
Note there are only 6 fitted values, so the residual versus fit plot will consist of only 6 stacks of points. The stacks should be similar in vertical dispersion.

> One of the assumptions for the GLM, in addition to normal residuals, is that the fitted and residuals values are not associated. In mathematical terms:
>
> V = Model + Res
>
> Var(V) = Var(Model + Res)
>
> Var(V) = Var(Model) + Var(Res) + Cov(Model,Res)
>
> If Cov(Model,Res) = 0, then we can partition Var(V)
>
> Var(V) = Var(Model) + Var(Res)
>
> If Cov(Model,Res) ≠ 0, then partioning of Var(V) cannot be trusted.

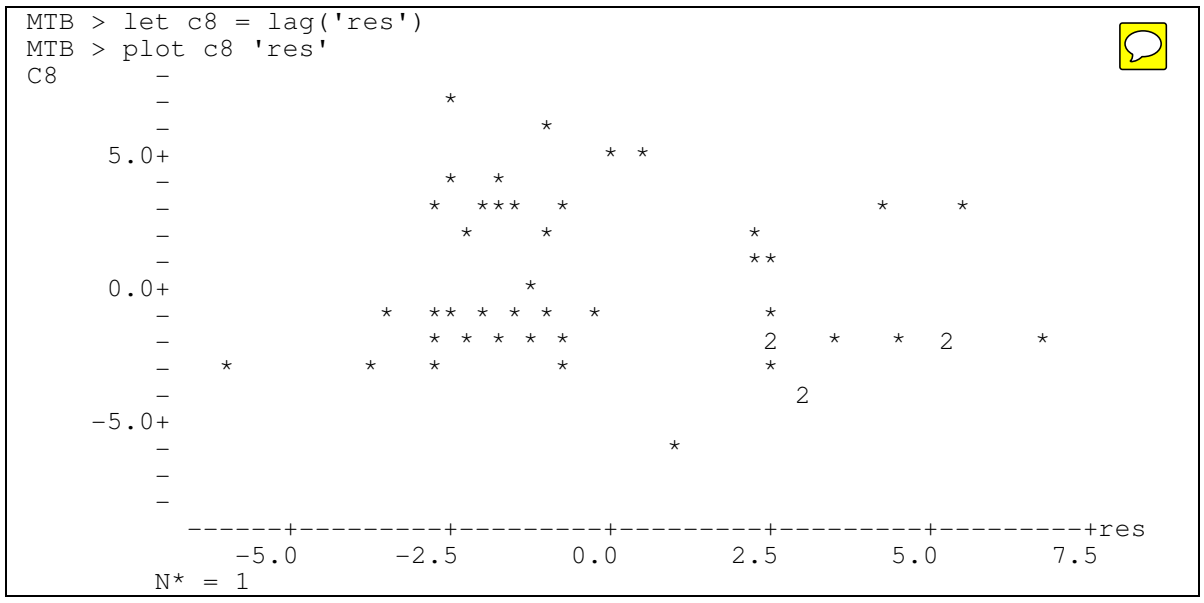
3. Evaluate model.

c. If n small, evaluate remaining assumptions for computing p-values.

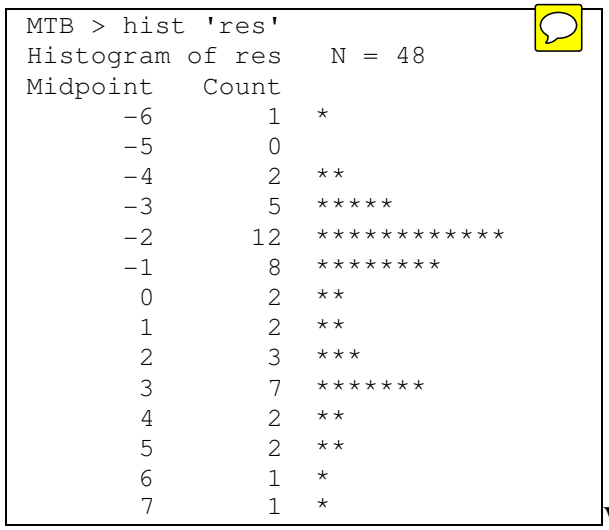
n = 48 so only large violations will distort p-values or confidence limits.

Sum(res) = 0? Yes
Independent? Yes

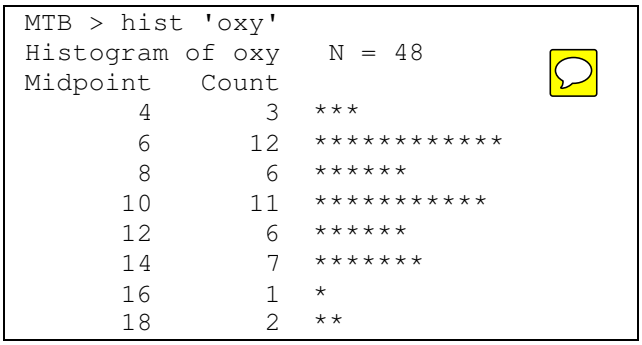
Normal ? Clearly bimodal. It is interesting that data do not show such strong deviation



from normal distribution.



```
MTB > hist 'res'
Histogram of res   N = 48
Midpoint   Count
      -6       1  *
      -5       0
      -4       2  **
      -3       5  *****
      -2      12  ************
      -1       8  ********
       0       2  **
       1       2  **
       2       3  ***
       3       7  *******
       4       2  **
       5       2  **
       6       1  *
       7       1  *
```



```
MTB > hist 'oxy'
Histogram of oxy   N = 48
Midpoint   Count
       4       3  ***
       6      12  ************
       8       6  ******
      10      11  ***********
      12       6  ******
      14       7  *******
      16       1  *
      18       2  **
```

Will this distort the estimate of p-value from F-distribution? As a rule of thumb, the p-value is unlikely to change by more than a factor of 2 if n > 30. In this analysis, n>30 so we will use the p-value from the cumulative distribution function, then compare this to the p-value by randomization.

4. State population and whether sample is representative.

When we draw conclusions from this sample, what is the population we are prepared to discuss ? The salinities were chosen by experimental design. We will view salinity as a fixed factor and hence infer only to these three levels of salinity. We will not take these

salinity levels as samples from all possible salinities. Hence we will not infer to any set of three salinities. In this example we will treat species as a fixed factor as well. That is, we will not infer to any pair of *Acmaea* species. We will infer only to the difference between these two species. Can we infer our results to these two species? To do so, we have to assume that the limpets in this study were representative of individuals from the entire population of *Acmaea scabra* and *A. digitalis*. However, we don't know if a representative sample was taken from the entire population of both species. It is unlikely that the limpets were collected in such a way that all limpets had an equal chance of being selected. There is not enough information to define the chance of being selected.

We can safely infer to a population of all possible measurements of O_2 consumption on in these two species, given the mode of collection. We will be considering a hypothetical population of possible measurements, not an enumerable biological population.

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

Research question is binary: does salinity and species affect oxygen consumption?
So hypothesis testing is appropriate (step 6).

6. State H_A H_o pairs, test statistic, distribution, tolerance for Type I error.

Analysis will focus first on the interaction term $\beta_{sp\ x\ sal} \cdot X_{sp} \cdot X_{sal}$
If the factors have interactive effects on the response variable, then the observed difference in O_2 consumption due to one factor (between the two groups of animal) will depend on the other factor (salinity). If there are significant interactive effects then the differences among species cannot be interpreted unless we know the salinity level.

The symbol $\beta_{sp\ x\ sal}$ stands for two parameters, which quantify the degree to which the effects of salinity in respiration depends on species.
These two parameters can be visualized in terms of dimensionless ratios.

Graph shows these ratios, as slopes of three lines.

$R_{50\%}$ = Mean($_{sal=50,\ sp=Ad}$) / Mean($_{sal=50,\ sp=As}$) = 12.326 / 12.174 = 1.013
$R_{75\%}$ = Mean($_{sal=75,\ sp=Ad}$) / Mean($_{sal=75,\ sp=As}$) = 7.338 / 7.890 = 0.93
$R_{100\%}$ = Mean($_{sal=100,\ sp=Ad}$) / Mean($_{sal=100,\ sp=As}$) = 7.429 / 10.561 = 0.703

$R_{75\%} / R_{50\%} = 0.93 / 1.013 = 0.92$	Ratios are similar (nearly equal to unity)
$R_{100\%} / R_{50\%} = 0.703 / 1.013 = 0.69$	Ratios are not similar $R_{100\%} < R_{50\%}$ Greater change at 100% than at 50%

Hypotheses for the interaction term.
The research hypothesis H_A is that $\beta_{sp\ x\ sal} = 0$

H_A: $\beta_{sp\ x\ sal} \neq 0$
H_o: $\beta_{sp\ x\ sal} = 0$

6. State H_A H_o pairs, test statistic, distribution, tolerance for Type I error.

Are there more specific hypotheses about the interaction term?

No, because there is no information from experiment on which to base directional hypotheses.

If the parameter values are not zero, then there will be variance.

Model at top of board on left.
ANOVA table at top of board on right.

The H_A / H_o pairs equivalent to those listed above are:

H_A: Var($\beta_{sp\ x\ sal}$) >0 or equivalently H_A: Var($\beta_{sp\ x\ sal} \cdot X_{sp} \cdot X_{sal}$) > 0

H_o: Var($\beta_{sp\ x\ sal}$) = 0 or equivalently H_o: Var($\beta_{sp\ x\ sal} \cdot X_{sp} \cdot X_{sal}$) = 0

If the interaction term is not significant, then research hypotheses concerning each of the other terms in the model become of interest because we can interpret the effects of on factor (such as salinity) regardless of the effects of the other factor (species).

- Hypotheses for the species term.

H_A: PopMean($_{sp=As}$) ≠PopMean($_{sp=Ad}$) The population means differ

H_o: PopMean($_{sp=As}$) = PopMean($_{sp=Ad}$) The population means do not differ

These hypotheses are equivalent to

H_A: $\beta_{Sp=Ad} \neq 0$

H_o: $\beta_{Sp=Ad} = 0$

The are also equivalent to following H_A / H_o for parameters.

H_A: Var(β_{sp}) > 0 There is variance present, due to species

H_o: Var(β_{sp}) = 0 The is no variance due to species.

Are there more specific hypotheses about parameters ? No

- Hypotheses for the salinity term. Fixed effects term so the contrast in means will be of interest.

H_A: PopMean($_{Sal=50\%}$) ≠PopMean($_{Sal=75\%}$)≠PopMean($_{Sal=100\%}$)

The population means differ among salinities.

H_o: PopMean($_{Sal=50\%}$) = PopMean($_{Sal=75\%}$) = PopMean($_{Sal=100\%}$)

The population means do not differ

These hypotheses are equivalent to

H_A: $\beta_{Sal} \neq 0$

H_o: $\beta_{Sal} = 0$

The H_A H_o pair above is equivalent to the following hypotheses.

H_A: Var(β_{sal}) > 0 There is variance present, due to salinity.

H_o: Var(β_{sal}) = 0 There is no variance present, due to salinity.

- Additional hypotheses for parameters in the source term ? Yes

H_A: $_{100\%} \neq (1/2)(_{75\%} + _{50\%})$ Means at reduced salinity differ from unreduced.

H_o: $_{100\%} = (1/2)(_{75\%} + _{50\%})$

State test statistic F-ratio

Distribution of test statistic F-distribution

Tolerance for Type I error 5% (conventional level)

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

Take df from beneath each term and place in table.

GLM	=	β_o	+	$\beta_{sp} \cdot X_{sp}$	+	$\beta_{sal} \cdot X_{sal}$	+	$\beta_{sp*sal} \cdot X_{sp} \cdot X_{Sal}$	+	ϵ
Source Total	=			Sp	+	Sal	+	Sp·Sal	+	Resid

Compute total degrees of freedom $df_{total} = n - 1 = 48 - 1 = 47$

Partition df_{total} according to model, using rules

Source	df	SS	MS	F	---->	p
Sp	1					
Sal	2					
Sp·Sal	2					
Res	?	? = 47 − 1 − 2 − 2 = 42				
Total	47					

two species, hence 2-1 = 1 df $df_{Sp} = 2 - 1 = 1$

three salinities, hence 3-1 = 2 df $df_{Sal} = 3 - 1 = 2$

$df_{Sp*Sal} = df_{Sp} \cdot df_{Sal}$ $df_{Sp*Sal} = 1 \cdot 2 = 2$

$df_{res} = df_{total} - df_{Sp} - df_{Sal} - df_{Sp*Sal}$ $df_{total} = 47 - 1 - 2 - 2 = 42$

7. ANOVA - Calculate then partition variance according to model.

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

By hand

$SS_{tot} = \Sigma^2 - n^{-1}(\Sigma)^2 = 5065.153 - 48^{-1} \cdot 461.74^2 = 623.4066$

$SS_{tot} = (n-1) * Var() = 47 * 13.26 = 623.407$

By Minitab

```
MTB> let k1=ssq('oxy')
MTB> print k1
```

Partition SS_{tot} using Minitab commands

Line up the GLM model with the ANOVA and GLM commands

$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_{1x2} X_{1x2} + \epsilon$

```
MTB> ANOVA   'Response' =   'X1'     'X2'     'X1'*'X2'
MTB> ANOVA     c1      =   c2       c3       c2*c3
MTB> GLM       c1      =   c2       c3       c2*c3
MTB> GLM       'oxy'        'sp'     'sal'    'sp'*'sal'
```

Both GLM and ANOVA commands will partition the variance according to the model

Y	=	β_o	+	$\beta_{sp} X_{sp}$	+	$\beta_{sal} X_{sal}$	+	$\beta_{sp*sal} X_{sp*sal}$	+	ϵ
623.41	=			16.64	+	181.32	+	23.93	+	401.52

(which is that of a two-way ANOVA).. The GLM command is more flexible, it allows unequal replication within the cells (in this example replication is equal: 8 observations in each of 6 cells). If we had a missing value (only 47 observations) we could not use the ANOVA command, but we could use the GLM command.

7. ANOVA - Partition df, SS, and construct ANOVA tablel.

Bring SS components from beneath model to table, to bring out relation of model to table.
Start with SS_{tot} at bottom, then add partitioned components SS_{sp} SS_{sal} $SS_{sp\ x\ sal}$ SS_{res}

Source	df	SS	MS	F	----> p
Sp	1	16.64			
Sal	2	181.32			
Sp·Sal	2	23.93			
Res	42	401.52			
Total	47	623.41			

r^2 = explained variance = SS_{model} / SS_{tot} = (623.41−401.52) / 623.41 = 36%
SS_{model} / SS_{tot} = coefficient of determination (for any GLM)
1 − coefficient of determination = coefficient of non-determination = 64%

Compute and table MS

Add to table

$MS_{Sp} = SS_{Sp} / df_{Sp}$ = 16.64
$MS_{Sal} = SS_{Sal} / df_{Sal}$ = 90.661
$MS_{Sp * Sal} = SS_{Sp * Sal} / df_{sp\ x\ sal}$ = 11.963
$MS_{res} = SS_{res} / df_{res}$ = 9.56

Compute and table F

Fixed effects for salinity and species, so all variance ratios taken relative to MS_{res}
$F = MS_{sp} / MS_{res} = 16.638 / 9.56 = 1.74$
$F = MS_{sal} / MS_{res} = 90.661 / 9.56 = 9.48$
$F = MS_{sp\ x\ sal} / MS_{res} = 11.963 / 9.56 = 1.2$

Add to table

Calculate Type I error from F-distribution.

$F_{2,42} = 1.251$	$p = 0.297$	interaction
$F_{1,42} = 1.74$	$p = 0.194$	species effect
$F_{2,42} = 9.483$	$p = 0.0004$	salinity effect

Draw picture of computational flow.
Add p-values to table

Statistical packages perform step 7 automatically: partition df, partition SS, compute MS, compute F, compute p-value from F-statistic, and produce ANOVA table.

8. Decide whether to recompute p-value.

The residuals are strongly bimodal, not normal.

n > 30 and as well, p is not near α (differs by more than a factor of 2)

Hence we will use the cumulative distribution function to compute p-values.

However, we will check our judgement by computing the randomized p-value.

To obtain randomization p-value, set up a control file and generate distributions based on hundreds of randomizations.

execute ANOVA
run 1000 randomizations
print results.

```
sample 48 'oxy' c7
unstack c7 c31 c32 c33;
subscripts 'sal'.
let k31 = mean(c31)
let k32 = mean(c32)
let k33 = mean(c33)
set c8
(k31 k32 k33)16
end
unstack c7 c34 c35;
subscripts 'sp'.
let k34 = mean(c34)
let k35 = mean(c35)
set c9
(k34 k35)24
end
let k1 = stdev(c7)*stdev(c7)*47
let k2 = stdev('sp')*stdev('sp')*47
let k3 = stdev('sal')*stdev('sal')*47
let k4 = stdev('fits')*stdev('fits')*47
let k5 = stdev('res')*stdev('res')*47
let k8 = stdev(c8)*stdev(c8)*47
let k9 = stdev(c9)*stdev(c9)*47
let k10 = k4 - k8 - k9
let k15 = (k8/k5)*(42/2)                # F sal
let k16 = (k9/k5)*(42/1)                # F sp
let k17 = (k10/k5)*(42/2)               # F sal*sp
stack c15 k15 c15
stack c16 k16 c16
stack c17 k17 c17
```

$F_{2,42}$ = 1.251 p = 386/1000 = 0.39 interaction term
$F_{1,42}$ = 1.74 p = 250/1000 = 0.25 species effect
$F_{2,42}$ = 9.483 p = 1/1000 = 0.001 salinity effect

How much did the p-values change?

interaction	F = 1.25	p_{ran} = 0.39	p_{cdf} = 0.297	0.39 / 0.297 = 1.3
species	F = 1.74	p_{ran} = 0.25	p_{cdf} = 0.19	0.25 / 0.19 = 1.3
salinity	F = 9.48	p_{ran} = 0.001	(poor estimate, only 1000 randomizations)	

9. Declare decision about terms.

Start with interaction term.
$p = 0.39 > \alpha = 5\%$ accept H_o: no interaction term
Interaction nowhere close to significant so proceed to main effects.
If interactive effects are present, then we do not proceed to main effects, see in the example in the next chapter.

Remove interaction term if not significant ?
We begin by looking at effect on MS_{res}

Source	df	SS	MS	F ----> p
Sp	1	16.64		
Sal	2	181.32		
Sp·Sal	2	23.93		
Res	42	401.52	9.56	
Total	47	623.41		

The interaction SS and df drop down into the residual if we omit interaction term.
The new MS_{res} is easily calculated.
$MS_{res} = (23.93 + 401.52) / (2 + 42) = 9.67$

In this case the MS_{res} changes only slightly, so the F-ratios will not change.
In general, there will be no gain (ie smaller error) unless the interaction MS is much smaller than the MS error. If MS error is not going to change, there is little reason to undertake a new analysis.

Proceed to main effects - species and salinity.
$p = 0.25 > \alpha = 5\%$ accept H_o: no difference in respiration between species

$p < 0.001 < \alpha = 5\%$ reject H_o accept $H_{A:}$ respiration depends on salinity

This two-way ANOVA is also known as a crossed design. One factor has been crossed with another. There are values of Y for each level of salinity within both species. In a crossed design we generally want to interpret the main effects. However, we cannot do this if the interaction term is significant. If the interaction term is significant, then it is a good idea to go back to the biology of the situation and reformulate the question so that the interaction term is of central importance, rather than the main effects.

10. Report and interpret parameters of biological interest.

The interaction term was not significant, so we conclude that the ratios (one species relative to another) are the same across the 3 salinities.
Oxygen consumption changes with salinity, in both species, in the same way.
$R_{50\%} = 1.013$
$R_{75\%} = 0.93$
$R_{100\%} = 0.703$
The design was capable of identifying a change in ratio of 0.7 to 1.0 as statistically significant.
(Strictly speaking the significance test is on differences, not on ratios. However, if the difference between two means is significant then by extension their ratio is greater than 1).
The species term was not significant, was not considered of biological interest in this experiment, and so is not examined.

The salinity term was of biological significance. Effects were statistically significant, so we conclude that respiration depends on salinity.

At this point we could do planned comparisons .
A logical set of comparisons would be 100% versus 70% then 100% versus 50%.
We can undertake two tests.
Equivalently, we can examine the means and standard errors because as an approximation, two means differ significantly if the difference is greater than two standard errors.

$\hat{\beta}_{sal=50} = 12.25\ \mu l\ min^{-1}\ mg^{-1}$ st.err = $0.800\ \mu l\ min^{-1}\ mg^{-1}$
$\hat{\beta}_{sal=75} = 7.61\ \mu l\ min^{-1}\ mg^{-1}$ st.err = $0.669\ \mu l\ min^{-1}\ mg^{-1}$
$\hat{\beta}_{sal=100} = 8.99\ \mu l\ min^{-1}\ mg^{-1}$ st.err = $0.868\ \mu l\ min^{-1}\ mg^{-1}$

We can see that reduction to 50% salinity increased respiration, while reduction to 75% did not.
We can draw this conclusion from inspection of the means and standard errors.

100% versus 75%
The difference is $8.99 - 7.61 = 1.38\ \mu l\ min^{-1}\ mg^{-1}$.
Two standard errors are $0.669 + 0.868 = 1.54\ \mu l\ min^{-1}\ mg^{-1}$.
100% versus 50%
The difference is $8.99 - 12.25 = -3.26\ \mu l\ min^{-1}\ mg^{-1}$.
Two standard errors are $0.669 + 0.800 = 1.47\ \mu l\ min^{-1}\ mg^{-1}$.

This is an approximation, not an exact test, but it is quick and easy.

Report conclusions based on parameters and some measure of uncertainty.
Based on the means and standard errors, we conclude that a drop in salinity to 50% increases respiration, while a drop in salinity to 75% does not change respiration.
Report means with standard errors as above.