

Model Based Statistics in Biology.

Part IV. The General Linear Model. Multiple Explanatory Variables.

Chapter 13.2 Fixed Effects ANOVA (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)

SC16_6_1.xls Ch13.xls

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)

ReCap (Ch 13) GLM with more than one categorical variable (ANOVA).

New concept, the interaction term.

Today: Two-way ANOVA.

One response variable Y as a function of two explanatory variables X_1 X_2 .
--

Both explanatory variables are categorical, <i>i.e.</i> on a nominal scale.

Interaction term not negligible.

Wrap-up. General Linear Model with two fixed factors.

The interaction term in the model was not significant at fixed 5% level.

However, there is evidence for an interactive effect ($LR = 18$).

The effects of one factor differ across the other factor.

When interactive effects are present we analyze one factor within each level of another factor

Background.

The example comes from a widely used and influential text by Snedecor and Cochran. George W. Snedecor taught statistics from 1913 to 1958 at the University of Iowa, where he was director of the statistical laboratory. William G. Cochran authored two other influential texts *Sampling Techniques* (3rd edition 1977) and *Experimental Designs* (with Gertrude M Cox) 2nd edition 1957.

In this example the authors judge that the interactive effect cannot be ignored, based on their extensive experience with experimental data.

The 8th edition of Snedecor and Cochran (1989) reported gains in weight in rats under six diets. Data were displayed on page 304 as Table 16.6.1, shown here.

	High Protein			Low Protein		
	Beef	Cereal	Pork	Beef	Cereal	Pork
	73	98	94	90	107	49
	102	74	79	76	95	82
	118	56	96	90	97	73
	104	111	98	64	80	86
	81	95	102	86	98	81
	107	88	102	51	74	97
	100	82	108	72	74	106
	87	77	91	90	67	70
	117	86	120	95	89	61
	<u>111</u>	<u>92</u>	<u>105</u>	<u>78</u>	<u>58</u>	<u>82</u>
Total	1000	859	995	792	839	787

1. Construct model

Data are: weight gains (grams) in rats fed 6 diets classified by source of protein (cereal, beef, or pork) and by level of protein (low or high). Ten rats/diet

Does weight gain depend on diet ?

Response variable

Weight gain: ΔM in grams (ratio scale)

Explanatory variables are protein source and protein level.

Source: X_S = source of protein in three categories: cereal, beef, or pork

Level: X_L = protein level in two categories, high or low.

Both explanatory variables are fixed factors.

Verbal model.

Weight gain depends on protein source and level.

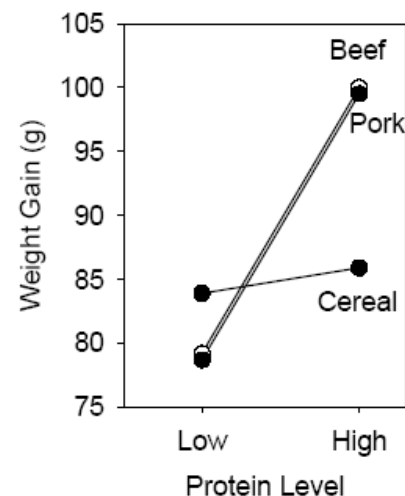
Graphical model.

Y-axis = Weight gain

X-axis = Low or high levels of protein

Connect two means of each of three sources

Graph suggests that growth depends on level for beef in pork more than for cereal.



Formal Model

Write GLM: $\Delta M = \beta_o + \beta_S \cdot X_S + \beta_L \cdot X_L + \beta_{S \times L} \cdot X_S \cdot X_L + \text{residual}$

S&R95 $V_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$

The model has been written in two forms. One is typical notation for the GLM, the other is from the tradition of experimental design (e.g. Snedecor and Cochran 1989, Sokal and Rohlf 1995) in which fixed factors are represented by greek letters.

2. Execute analysis.

Place data in model format:

Column labelled M for mass. This is the response variable weight gain.

Column labelled XS with explanatory variable, $X_S \in \{-1, 0, 1\}$

-1 = cereal, 0=beef, 1 = pork

Column labelled XL with explanatory variable, $X_L = 0$ (low) or 1 (high)

These are labels (categories), not numbers on ratio scale.

The model has three terms, one for each factor and one for the interactive effect. The interaction term measures the degree to which weight gain by source of protein depends on level of protein.

Code model statement in statistical package according to the GLM

$$\Delta M - \beta_0 = \beta_S \cdot X_S + \beta_L \cdot X_L + \beta_{S \times L} \cdot X_S \cdot X_L + \varepsilon$$

```
MTB> GLM 'M' = 'XS' 'XL' 'XS'*'XL'  
SUBC> fits c4;  
SUBC> res c5.
```

If you are using code in a stat package write the code in the space below.

The grand mean: $\hat{\beta}_0 = 87.9$ g

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

	Low	High
Cereal	83.9	85.9
Beef	79.2	100
Pork	78.7	99.5

These values are used to compute a residual value, one for each of 60 observations.

Residuals can also be calculated from the 6 parameter estimates produced the GLM command. Here are the GLM parameters produced by the SPlus statistical package.

	Value	Std.Error	t value	Pr(> t)
(Intercept)	87.867	1.891	46.4654	0.000
Level	7.2667	1.891	3.8427	0.000
Source1	2.35	2.316	1.0147	0.315
Source2	0.6167	1.3371	0.4612	0.647
LevelSource1	4.7	2.316	2.0294	0.047
LevelSource2	1.5667	1.3371	1.1716	0.247

3. Evaluate model.

Plot residual versus fitted values.

There are six stacks of values, one stack for each of six cell means (fitted values).

Straight line model.

No line fitted in model

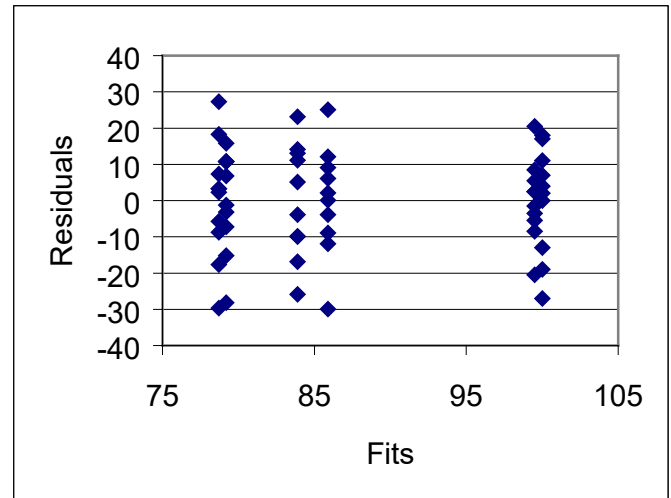
Error model. $n = 60$ so only large violations will distort estimates.

Homogeneous residuals?

No systematic change in residuals with increase in fitted values (*i.e.* no fans or cones).

The six stacks in the plot are similar in vertical dispersion.

The residuals are judged homogeneous.

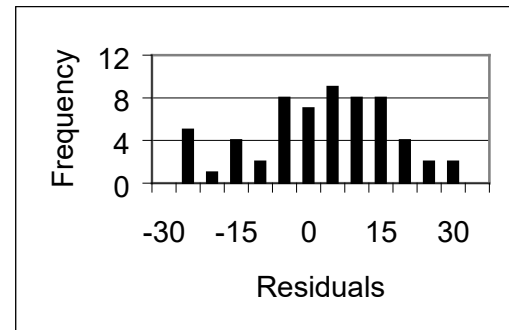


Error model.

Normal residuals ?

The residuals are skewed to the left side, too many large negative residual values.

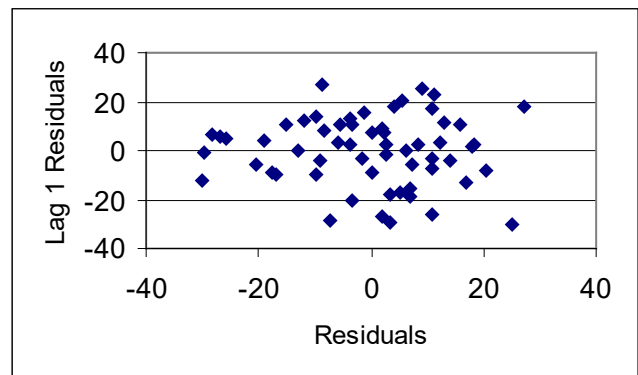
The residuals deviate from normal.



Independent?

Graph shows no evidence of upward or downward trend. The residuals are judged independent.

Summary. The residuals are homogeneous and independent, but they deviate from normal. However, because $n=60$, we judge that parameter estimates and p-values will be little affected.



4. Partition df and SS according to model.

Compute total degrees of freedom

$$dftotal = n - 1 = 60 - 1 = 59$$

Partition $dftotal$ according to model, using rules

two levels, hence $2 - 1 = 1$ df

$$dfL = 2 - 1 = 1$$

three sources, hence $3 - 1 = 2$ df

$$dfS = 3 - 1 = 2$$

$dfL * S = dfL - dfS$

$$dfSp * Sal = 1 * 2 = 2$$

$dfres = dftotal - dfSp - dfSal - dfSp * Sal$

$$dfres = 59 - 1 - 2 - 2 = 54$$

Y / β_o	=	$\beta_S X_S$	+	$\beta_L X_L$	+	$\beta_{S*L} X_S * L$	+	,
59	=	1	+	2	+	2	+	54
16198.93	=	266.5	+	3168.3	+	1178.2	+	11585.7
Full model								Reduced model

4. Calculate LR

$$1 - R^2 = \text{Full/Reduced} = (11585.7/16198.93) = 71.5\%$$

$$LR = (0.715)^{-60/2} = 23277$$

The research hypothesis is far more likely than a simple one-parameter model, the mean. There is very strong evidence for the omnibus model with three terms.

Continue with analysis of individual terms

5. State population and whether sample is representative.

The population is not enumerable (number of individuals, spatial units, *etc.*). The population to which we infer is a large number of repetitions of the experiment based on the experimental protocol, including the procedural statement for measurement of variables. The protein levels and sources were chosen by experimental design. We will view these as fixed factors and hence infer only to these three sources and two levels of protein. We assume that rats were assigned randomly to treatments, so that the results are representative of any other experiment with the same design. We will be considering a hypothetical population of repetitions of the experiment.

5. Choose mode of inference. Is hypothesis testing appropriate?

Evidential inference is appropriate where the costs of Type I versus Type II error are unknown. Evidential inference rests on a measure of evidence, not on a measure of uncertainty. We will use evidential inference then compare it to the compare practice of decision theoretic inference against a fixed Type I error.

6. List specific hypothesis pairs.

Analysis will focus first on the interaction term $\beta_{S \times L} \cdot X_S \cdot X_L$

If the factors have interactive effects on the response variable, then the observed difference in weight gain due to one factor (protein source) will depend on the other factor (protein level). If there are significant interactive effects then the weight gains among the three sources cannot be interpreted unless we know the protein level.

The symbol $\beta_{S \times L}$ stands for two parameters, which quantify the degree to which the effects of protein source on weight gain depends on level. One parameter measures the change in change in weight from low to high level in beef, relative to cereal. The second parameter measures the change in weight from low to high in pork, relative to beef.

6. List specific hypothesis pairs.

Hypotheses for the interaction term.

The research hypothesis H_A is that $\beta_{S \times L} = 0$

$$H_A: \beta_{S \times L} \neq 0 \quad H_0: \beta_{S \times L} = 0$$

Are there more specific hypotheses about the interaction term?

Yes. The experimenters expected the weight gain to be greater from high to low for animal protein sources (beef and pork) than for cereal.

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

The H_A / H_0 pairs equivalent to those listed above are:

$$H_A: \text{Var}(\beta_{S \times L}) > 0 \quad \text{or equivalently} \quad H_A: \text{Var}(\beta_{S \times L} \cdot X_S \cdot X_L) > 0$$

$$H_0: \text{Var}(\beta_{S \times L}) = 0 \quad \text{or equivalently} \quad H_0: \text{Var}(\beta_{S \times L} \cdot X_S \cdot X_L) = 0$$

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 one factor (*e.g.* protein source) regardless of the effects of the other factor (*e.g.* protein level).

Hypotheses for the protein level term.

Fixed effects term so the contrast in means will be of interest.

$$H_A: \text{PopMean}(\Delta M_{L=low}) < \text{PopMean}(\Delta M_{L=high}) \quad \text{The population means differ}$$

$$H_0: \text{PopMean}(\Delta M_{L=low}) = \text{PopMean}(\Delta M_{L=high}) \quad \text{The population means do not differ}$$

These hypotheses are equivalent to following H_A / H_0 for parameters.

$$H_A: \text{Var}(\beta_L) > 0 \quad \text{There is variance present, due to level}$$

$$H_0: \text{Var}(\beta_L) = 0 \quad \text{There is no variance due to level.}$$

Are there more specific hypotheses about parameters? Yes. We might expect that weight gain is greater for high level than low levels of protein.

Hypotheses for the protein source term. Fixed effects term so the contrast in means will be of interest.

$$H_A: \text{PopMean}(\Delta M_{S=cereal}) \neq \text{PopMean}(\Delta M_{S=beef}) \neq \text{PopMean}(\Delta M_{S=pork})$$

$$H_A: \text{PopMean}(\Delta M_{S=cereal}) \neq \text{PopMean}(\Delta M_{S=beef}) \neq \text{PopMean}(\Delta M_{S=pork})$$

The population means differ among protein source.

$$H_0: \text{PopMean}(\Delta M_{S=cereal}) = \text{PopMean}(\Delta M_{S=beef}) = \text{PopMean}(\Delta M_{S=pork})$$

The population means do not differ

These hypotheses are equivalent to

$$H_A: \beta_S \neq 0$$

$$H_0: \beta_S = 0$$

The H_A / H_0 pair above is equivalent to the following hypotheses.

$$H_A: \text{Var}(\beta_S) > 0 \quad \text{There is variance present, due to protein source.}$$

$$H_0: \text{Var}(\beta_S) = 0 \quad \text{There is no variance present, due to protein source.}$$

Additional hypotheses for parameters in the source term? Yes

$$H_A: \Delta M_{cereal} < ((1/2)(\Delta M_{beef} + \Delta M_{pork})) \quad \text{Growth rates for cereal less than those for animal sources of protein.}$$

$$H_0: \Delta M_{cereal} = ((1/2)(\Delta M_{beef} + \Delta M_{pork}))$$

6. State test statistic, distribution, Type I error.

Test statistic	F-ratio
Distribution of test statistic	F-distribution
Type I error	None for evidentialist inference. 5% for decision theoretic frequentist inference.

7. ANOVA table.

Place factors, df, SS in table.

Source	df	SS	MS	F	----> p
Source	2	266.5			
Level	1	3168.3			
Source x Level	2	1178.2			
Res	54	11585.7			
Total	59	16198.9			

7. ANOVA table – Evidentialist.

Calculate LR from SS for each term in the model.

Source	df	SS	SSresid = SSreduced	SSo = SSfull	SSo/SSres	LR	Evidence
Source	2	267	11586	11853	1.023	2.0	Inadequate
Level	1	3168	11586	14754	1.273	1411	Very strong
S * L	2	1178	11586	12764	1.102	18.3	Nearly adequate
Error	54	11586					
Total	59						

7. ANOVA table – Frequentist with fixed Type I error.

Calculate MS, F, and p

Source	df	SS	MS	F	p	
Source	2	266.5333	133.2667	0.62	0.541132	Can't reject H_0
Level	1	3168.267	3168.267	14.77	0.000322	Reject H_0
S * L	2	1178.133	589.0667	2.75	0.073188	Can't reject H_0
Error	54	11586.00	214.5556			
Total	59					

8. Check p-value by using randomization.

$30 < n < 100$, residuals homogeneous, some deviation from normal distribution of errors, but because the sample was large, we judge that randomization will not change the p-value judged substantially. We check that judgement.

	p-value	p-random	p-value/p-random
Source	0.541	$2688/5000 = 0.538$	1.006
Level	0.000322	$1/5000 = 0.0002$	1.56 (poor estimate)
Source x Level	0.0732	$351/5000 = 0.070$	1.04

Note that the p-value for Level was poorly estimated because only 5000 randomizations were run. The p-value from the F-distribution was 3 in 1000 (or 15 in 5000), so our estimate can be no better than 1 part in 15. If we had used 10,000 randomizations our estimate the p-value would improve to 1 part in 30.

Our judgement that randomization would have little effect was correct.

9. Report statistical conclusion.

Start with interaction term.

$$LR = 18 \quad p = 0.073$$

The interactive term 18 times more likely than not.

Snedecor and Cochran (1989 p305) comment as follows.

“In factorial designs “it often happens that a few comparisons comprising the main effects have substantial interactions, while the majority of the comparisons have negligible interactions.

Consequently, the F-test of the AB interaction sum of squares as a whole is not a good guide as to whether interactions can be ignored. It is wise to look over the two-way table of treatment totals or means before concluding that there are no interactions, particularly if the F is larger than 1.”

Here is the table of cell means and marginal means, followed by table of contrasts with the grand mean, and table of contrasts from high to low in each source of protein. To examine the interactive effects, we subtract the grand mean from each cell mean. Then we subtract the marginal means from each cell.

cell, row, column, and grand means			
	Low	High	mean
cereal	83.9	85.9	84.9
beef	79.2	100.0	89.6
pork	78.7	99.5	89.1
mean	80.6	95.1	87.9
subtract grand mean			
	Low	High	mean
cereal	-3.967	-1.967	-2.967
beef	-8.667	12.133	1.733
pork	-9.167	11.633	1.233
mean	-7.267	7.267	
subtract grand mean and marginal means			
	Low	High	mean
cereal	6.267	-6.267	0.000
beef	-3.133	3.133	0.000
pork	-3.133	3.133	0.000
mean	0.000	0.000	

The contrast from low to high in cereal (6.267 g) is twice the magnitude of the contrast in beef and pork (−3.133) and opposite in sign. The contrast is $6.3/85 = 7\%$ of the mean. There is weak evidence ($LR = 18$) for an interactive effect.

9. Form a decision about terms.

Instead of ignoring the interaction term because the p-value was just short of statistical significance at the 5% level, we consider the contrasts in addition to the means within each level of the other fixed factor. Following the advice of Snedecor and Cochran (1989) we use judgement to temper our conclusion about the presence of an interactive effect. We judge that interactive effects need to be considered in this analysis.

Interactive effects mean that we cannot interpret one main effect independently of the other. To interpret the results in light the possibility of interactive effects we break down the two-way table by comparing means across one factor within in each level of the other factor. We can examine the differences from low to high within each level of protein source, as in the table above. Alternatively, we could examine the differences across sources within each of the two levels.

10. Report science conclusion.

We judge that interactive effects could not be ignored in this experiment. We report contrasts across one factor (protein level) in each class of the other factor (protein source).

The results of this experiment are conveniently summarized as differences of means within each protein source, along with confidence limits on the difference.

				SS		MS=			
				Low	High	Sum(SS)/18	sterr		
cereal	Low	High	Diff	2220.9	2030.9	236.2	3.437	-5.22	9.22
beef	Low	High	Diff	1735.6	2062.0	211.0	3.248	13.98	27.62
pork	Low	High	Diff	2464.1	1072.5	196.5	3.134	14.22	27.38

The confidence limits for cereal include zero (no difference in growth at low and high levels of protein). The confidence limits for beef and pork exclude zero (exclude no growth). These results are consistent with the graphical display of the means, which suggested that growth depends on level for beef and pork more than for cereal.

The coefficients reported by the statistical package lead to a similar conclusion. Here are the results from the Minitab package, which match those shown earlier from SPlus.

Term	Coef	SE Coef	T	P	
Constant	87.867	1.891	46.47	0.000	$\hat{\beta}_0$
Level					
0	-7.267	1.891	-3.84	0.000	beef = 0
Source					
-1	-2.967	2.674	-1.11	0.272	cereal = -1
0	1.733	2.674	0.65	0.520	beef = 0
Level*Source					
0 -1	6.267	2.674	2.34	0.023	cereal = -1
0 0	-3.133	2.674	-1.17	0.246	beef = 0

The contrast from low to high level in cereal (6.267 g) is twice the magnitude of the contrast in beef and pork (-3.133) and opposite in sign. The contrast for cereal differs significantly from the other two ($p = 0.023 < \alpha = 5\%$).