

Model Based Statistics in Biology.

Part IV. The General Linear Mixed Model..

Chapter 13.3 Fixed x Random Effects (Paired t-test)

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)

Sleep data, two drugs
Ch13.xls

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)

Estimation is concerned with the specific value of an unknown population parameter.

Confidence limits use a fixed error rate to span an interval that contains the parameter.

Hypothesis testing uses the logic of the null hypothesis to declare a decision about a parameter

ReCap (Ch 9, 10,11) The General Linear Model - 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

Today: Paired t-test.

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

Both explanatory variables are categorical, on a nominal scale.

One categorical variable is fixed, the other is random.

This is called a mixed model

Wrap-up.

The paired t-test is a General Linear Model with two categorical variables, *i.e.* two explanatory variables on a nominal scale.

One of these is fixed (2 classes only), the other is random (many classes).

We are interested in the fixed effects, controlled for the random effects.

Adding the random factor to the model reduces the error variance, resulting in a more sensitive test. This is called statistical control.

Introduction. A common experimental design is to measure all levels of a fixed variable within each of several units that vary randomly. For example, we measure the effects of first one drug, then another, on each of 10 patients. The advantage of this design is that variation among units can be removed from the analysis, allowing a more sensitive test of the factor of interest. This design is called a Randomized Complete Block (RCB) design. If the fixed variable has only two classes, this design is called a paired t-test. To illustrate this analysis, we use data from Cushny and Peebles (1905), which Gossett used to introduce the t-test.

1. Construct model

Data are: hours of extra sleep with two drugs Hyoscyamine (Drug A) and L Hyoscine (Drug B), each administered to 10 subjects. Values reported are averages. The pairing across subject allows us to remove the effects of individual variation.

Subject	Drug A	Drug B
1	0.7	1.9
2	-1.6	0.8
3	-0.2	1.1
4	-1.2	0.1
5	-0.1	-0.1
6	3.4	4.4
7	3.7	5.5
8	0.8	1.6
9	0.0	4.6
10	2.0	3.4

Response variable. T = hours of extra sleep

Explanatory variables are drug and subject.

Drug X_D = [Drug A, Drug B]

Drug is a nominal scale variable with two categories.

It is a fixed effect.

Subject X_S = subject [1,2,3,etc.]. Ten categories, nominal scale.

This is a random effect, because the mean value for each subject varies randomly.

The levels are not chosen by the investigator.

We are inferring to a population of similar subjects.

Here is a table of variables and their characteristics.

Name	Variable			Response or Explanatory	Categorical or Covariate	Fixed or Random	Crossed or Nested
	Symbol	Units	Scale				
Extra sleep	T	hours	ratio	response	N/A	N/A	N/A
Drug	X_D	A or B	nominal	explanatory	categorical	fixed	N/A
Subject	X_S	#1 to #10	nominal	explanatory	categorical	random	N/A
Drug x Subj	$X_D \times X_S$	2 by 10	nominal	explanatory	categorical	mixed	crossed

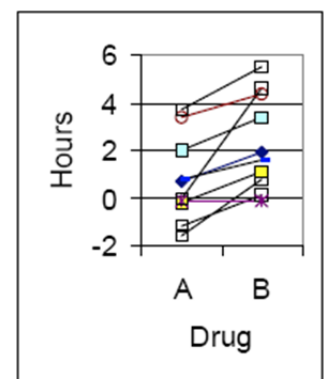
Verbal model. Hours of extra sleep depend on drug.

Graphical model.

Connect mean value at two levels for each subject.

Graph shows increase from Drug A to Drug B in all subjects.

There appears to be little or no interactive effect, because the lines rise from left to right, in a more or less parallel fashion.



1. Construct formal model

To begin, we write the formal model with the interaction term.

$$\begin{array}{l} \text{Write GLM:} \quad T = \beta_o + \beta_D X_D + \beta_S X_S + \beta_{D \times S} X_D X_S + \text{residual} \\ \text{Components:} \quad T_{ijk} = \mu + \alpha_i + S_j + (\alpha \cdot S)_{ij} + \epsilon_{ijk} \end{array}$$

The model is shown in two forms. Both have the same structure --three explanatory terms, one for drug (a fixed effect), one for subject (a random effect), and one for interactive effects--the dependence of drug effect on subject. The GLM notation shows model parameters and variables. The variance component notation assigns a greek letter (α) to the fixed factor and a roman letter (S) to the random factor. The interaction term is mixed. For future reference we can use these rules.

$$\begin{array}{l} \text{Random x Fixed} = \text{Mixed} \\ \text{Fixed x Fixed} = \text{Fixed} \\ \text{Random x Random} = \text{Random} \end{array}$$

The component notation shows subscripts: $i = 2$ levels of drug (A and B), $j = 10$ subjects, and $k =$ number of measurements per subject. In this example we have one measurement per subject at each level of treatment. We have $ijk = 2 \cdot 10 \cdot 1 = 20$ observations.

In the extra sleep data we have enough data to estimate a mean difference between drugs and a mean for each subject but we do not have enough data to estimate the mixed (interactive) term. If we estimate the interaction term there will be no degrees of freedom for the error term. The interaction term is not a fixed effect, so we can let it drop into the residual by omitting it from the model.

$$\begin{array}{l} \text{GLM:} \quad T = \beta_o + \beta_D X_D + \beta_S X_S + \epsilon \\ \text{Mixed model} \quad T_{ij} = \mu + \alpha_i + S_j + \epsilon_{ij} \end{array}$$

The result is a new residual term ($\epsilon_{ij} = (\alpha \cdot S)_{ij} + \epsilon$) that consists of a mixed and random term.

2. Execute analysis.

Place data in model format:

- Column labelled T, with response variable hours of extra sleep
 - Column labelled X_D with explanatory variable, $X_D = 0$ or 1
 - Column labelled X_S with explanatory variable, $X_S = 0, 1 \dots 10$
- These are labels (categories), not numbers on ratio scale.

Use the model to execute the analysis in a statistical package.

$$T = \beta_o + \beta_S X_S + \beta_D X_D + \epsilon$$

2. Execute analysis.

Here are the calculations in spreadsheet format.

The grand mean.

$$\beta_0 = 20^{-1} \sum T = 20^{-1} \cdot 30.8 = 1.54 \text{ hours}$$

The fitted values are computed from the row and column means.

$$\text{mean}(T_{D=A}) = 0.75 \text{ hours}$$

$$\hat{\beta}_{D=A} = (0.75 - 1.54) = -0.79 \text{ hours}$$

$$\text{mean}(T_{D=B}) = 2.33 \text{ hours}$$

$$\hat{\beta}_{D=B} = (2.33 - 1.54) = +0.79 \text{ hours}$$

$$\text{mean}(T_{S=1}) = 1.30 \text{ hours}$$

$$\hat{\beta}_{S=1} = (1.30 - 1.54) = -0.24 \text{ hours}$$

$$\text{mean}(T_{S=2}) = -0.40 \text{ hours}$$

$$\hat{\beta}_{S=2} = (-0.40 - 1.54) = -1.94 \text{ hours}$$

etc.

		Drug	Subject		
Hrs		Effect	Effect	fits	res
T	β_0	β_D	β_S	$\beta_0 + \beta_D + \beta_S$	
0.7	1.54	-0.79	-0.24	0.51	0.19
-1.6	1.54	-0.79	-1.94	-1.19	-0.41
-0.2	1.54	-0.79	-1.09	-0.34	0.14
-1.2	1.54	-0.79	-2.09	-1.34	0.14
-0.1	1.54	-0.79	-1.64	-0.89	0.79
3.4	1.54	-0.79	2.36	3.11	0.29
3.7	1.54	-0.79	3.06	3.81	-0.11
0.8	1.54	-0.79	-0.34	0.41	0.39
0.0	1.54	-0.79	0.76	1.51	-1.51
2.0	1.54	-0.79	1.16	1.91	0.09
1.9	1.54	0.79	-0.24	2.09	-0.19
0.8	1.54	0.79	-1.94	0.39	0.41
1.1	1.54	0.79	-1.09	1.24	-0.14
0.1	1.54	0.79	-2.09	0.24	-0.14
-0.1	1.54	0.79	-1.64	0.69	-0.79
4.4	1.54	0.79	2.36	4.69	-0.29
5.5	1.54	0.79	3.06	5.39	0.11
1.6	1.54	0.79	-0.34	1.99	-0.39
4.6	1.54	0.79	0.76	3.09	1.51
3.4	1.54	0.79	1.16	3.49	-0.09

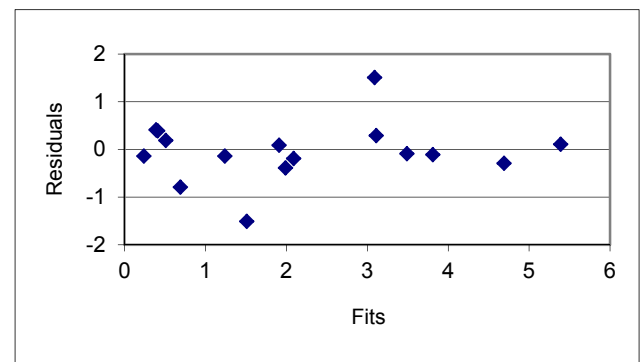
3. Evaluate the model.

Structural model.

No regression line, so skip this evaluation.

Error model.

Homogeneous? Vertical dispersion appears to decrease slightly from left to right. However, we have little information (very few points) on the far right side.

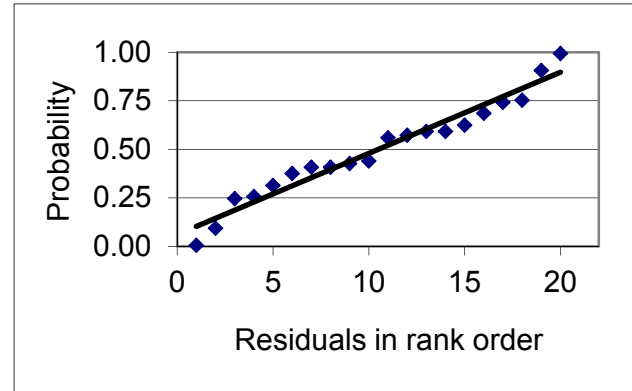


3. Evaluate the error model.

Error model

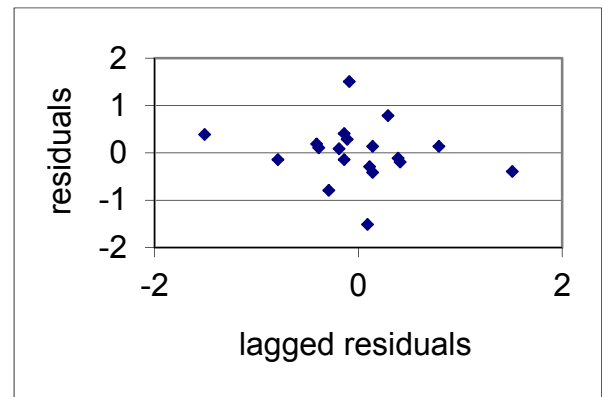
Normal ?

The residuals are normal, they closely follow the trend line in the normal probability plot.

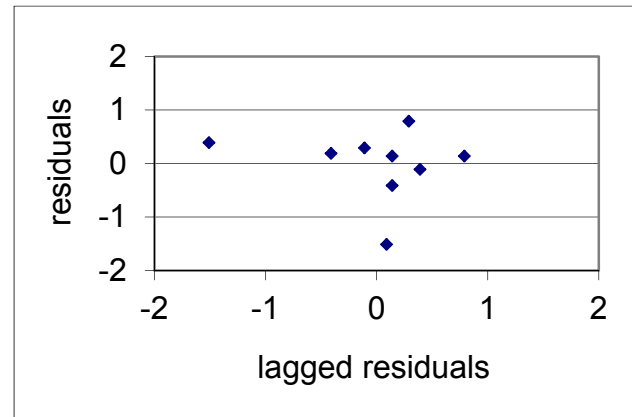


B3. Independent? Graph suggests some pattern of negative correlation. There are two mirror image patterns of 4 dots each in the graph.

Looking at the residuals, we see that for every value within Drug A, there is a value of the same magnitude and opposite sign within Drug B.



If we plot only the residuals from Drug A, the pattern disappears. We see that for factorial designs, we can expect patterns to appear when there are only two values per category. We expect patterns such as this to appear when we are fitting many parameters, relative to the number of observations. In this case we have 11 parameters for 20 observations.



4. State population and whether sample is representative.

When we draw conclusions from this sample, what is the population to which we are inferring? We are inferring only to these two drugs. Drug is a fixed factor because we are interested in only these two drugs. The subjects belong to a larger population, other subjects. We are inferring to this population, so we take Subject as a random effect. We do not know if the subjects are representative of all people or merely to people similar in age and medical history. We would require a well executed clinical trial with randomly chosen subjects from a population with known characteristics (age, sex, medical history) before making strong claims about the efficacy of one drug relative to another.

4. Partition df and SS according to model

GLM	$T - \beta_o = \beta_S \cdot X_S + \beta_D \cdot X_D + \epsilon$
Source	Total = Subject + Drug + Resid
df	20 - 1 = 10 - 1 + 2 - 1 + 19 - 9 - 1
SS	77.37 = 58.08 + 12.48 + 6.81

Calculate likelihood ratio from SS.

$$R^2 = (58.08 + 12.48) / 77.37 = 91\% \quad 1 - R^2 = 8.8\%$$

$$LR = (0.088)^{(-20/2)} = 3.6 \times 10^9$$

The research hypothesis is for more likely than a simple one-parameter model, the mean. There is very strong evidence for the omnibus model with two terms. Most of the explained variance is due to subject. We continue to listing of the hypothesis pair for the fixed term in model.

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

Because there is a prospect of clinical use of a drug, we will compute Type I error (use of an ineffective drug) by limiting Type I error to fixed value of $\alpha = 5\%$.

6. State H_A H_o pairs, test statistic, distribution, and use of Type I error.

Hypothesis testing will focus on the drug effect. We are not comparing the response of the subjects. The subject factor is random and so could be dropped into the residual term. However this inflates rather than reduces the residual mean square. Retaining the subject factor in the model reduces the error term and results in a more sensitive test, one able to pick up smaller differences than a model lacking the subject factor.

Hypotheses for the drug term.

$H_A: \text{Mean}(T_{D=A}) \neq \text{Mean}(T_{D=B})$ The population means differ

$H_0: \text{Mean}(T_{D=A}) = \text{Mean}(T_{D=B})$ The population means do not differ

These hypotheses are equivalent to

$$H_A: \beta_D \neq 0$$

$$H_0: \beta_D = 0$$

Are there more specific hypotheses about parameters? No

State test statistic

F-ratio

Distribution of test statistic

F-distribution

Type I error held to fixed value

$\alpha = 5\%$ (conventional level)

7. ANOVA - Table Source, df, SS. Calculate MS, F, p

Here are the SS calculations in a spreadsheet, based on data equations.

Hours of extra sleep					
hrs	hrs	Drug A-B	fits	res	
Drug A	Drug B				
0.7	1.9	1.2	1.58	-0.38	
-1.6	0.8	2.4	1.58	0.82	
-0.2	1.1	1.3	1.58	-0.28	
-1.2	0.1	1.3	1.58	-0.28	
-0.1	-0.1	0.0	1.58	-1.58	
3.4	4.4	1.0	1.58	-0.58	
3.7	5.5	1.8	1.58	0.22	
0.8	1.6	0.8	1.58	-0.78	
0.0	4.6	4.6	1.58	3.02	
2.0	3.4	1.4	1.58	-0.18	
	Mean	1.58	1.58	0.00	
	SS	13.62	0.00	13.62	
	df	9	0	9	
	MS	1.51			
	t	4.06			
	p	0.0028			

$$t = \frac{1.58}{\sqrt{(13.62/9)/10}}$$

Note that $t^2 = 4.06^2 = 16.5$

Hours of extra sleep				Drug	Subject		
Subject	Drug-Ctl	Hrs		Effect	Effect	fits	res
	0=A,1=B	T	bo	bD	bS	bo+bD+bS	
1	0	0.7	1.54	-0.79	-0.24	0.51	0.19
2	0	-1.6	1.54	-0.79	-1.94	-1.19	-0.41
etc							
9	1	4.6	1.54	0.79	0.76	3.09	1.51
10	1	3.4	1.54	0.79	1.16	3.49	-0.09
	Mean	1.54	1.54	0	0	1.54	0
	SS	77.37		12.482	58.078	70.56	6.808
	df	19		1	9		9
	Source	df		SS	MS	F	p
	Subject	9		58.078	6.4531		
	Drug	1		12.482	12.482	16.50088	0.0028
	Residual	9		6.808	0.7564		
				77.368			

Here are the ANOVA table calculations in spreadsheet form.

We see that $F = t^2$ (shown above)

7. ANOVA - Calculate MS, F, p

Source	df	SS	MS	F	---->	p
Subject	9	58.078				
Drug	1	12.482	12.482	16.50		0.0028
<u>Res</u>	<u>9</u>	<u>6.808</u>	0.756			
Total	19	77.368				

Dropping the Subject term from the model, would add the SS and df to the residual. This increases the residual from $MS = 0.756$ to $MS = (58.078 + 6.808) / (9 + 9) = 3.604$. This in turn decreases the F-ratio to $F = 12.482 / 3.604 = 3.46$. It increases the estimate of Type I error to $p = 0.079$. The test statistic no longer meets the $\alpha = 5\%$ criterion. The power of the test is decreased when the among subject variability drops into the residual.

8. Recompute p-value?

We judge that the assumptions were met.

9. Report statistical conclusion.

Random factors are not tested. For the drug effect we report the F-ratio and p-value.

$$F_{1,9} = 16.5 \quad p = 0.0028 < \alpha = 5\%$$

We reject H_0 that the the observed difference is merely chance. We cannot, however, say that the model with the drug effect is better than the null model. To do this we would cite the evidence ratio $LR = 3.6 \times 10^9$ times better than the H_0 .

10. Report science conclusions and interpret parameters of biological interest.

No means are reported or compared for the subject term because it is a random factor and so the means are of no interest.

The confidence limits for the two drug groups overlap because the among subject variance is not controlled statistically.

	st.err	Lower CL	Upper CL
$\text{mean}(T_A) = 0.75 \text{ hours}$	0.5657	-0.53 hours	2.03 hours
$\text{mean}(T_B) = 2.33 \text{ hours}$	0.6332	+0.80 hours	3.06 hours

Here are the confidence limits for the average difference, controlled for among subject variation. Note that the standard error of the difference is smaller than the standard errors of each group. This is because the among subject variance has been removed.

	st.err	Lower CL	Upper CL
$\text{mean}(T_B - T_A) = 1.58 \text{ hours}$	0.3890	0.70 hours	2.46 hours

The confidence limits do not include zero; we can exclude the null hypothesis of no effect.

We conclude that the model with the drug term is more likely than the model with no drug term, provided we include subject, the random factor, in the model.