Model Based Statistics in Biology.
Part IV. The General Linear Model. Multiple Explanatory Variables.
Chapter 13.3 Fixed*Random Effects (Paired t-test)

Sleep data, two drugs
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).

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.

Wrap-up.

The paired t-test is a General Linear Model with two classification 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 very common design is to measure both levels of a fixed variable on the each of several units that vary randomly. For example, we measure 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. If the fixed variable has only two classes, this design is called a paired t-test. To illustrate this analysis, we return to the sleep data (Cushny and Peebles), which Gossett used to introduce the t-test.

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

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.

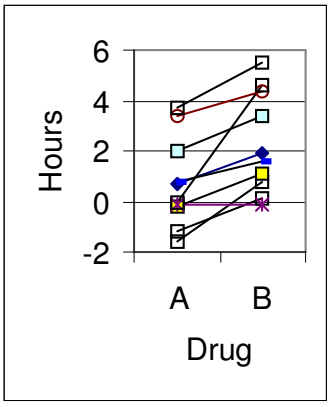
Response variable. T = hours of extra sleep

Explanatory variables are drug and subject.
Drug X_D = Drug A, Drug B Two categories, nominal scale. This 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 and is not under the control of the investigator.

Verbal model. Hours of extra sleep depends 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.

Formal Model

Write GLM: $T = \beta_o + \beta_D \cdot X_D + \beta_S \cdot X_S + \beta_{D\,x\,S} \cdot X_D \cdot X_S + \text{residual}$

S&R95 $T_{ijk} = \mu + \alpha_i + B_j + (\alpha \cdot B)_{ij} + \epsilon_{ijk}$

A model has been written using two forms of notation. One is typical notation for the GLM, the other typical in experimental design.

There are three explanatory terms, one for drug (fixed effect, greek symbol α), one for subject (random effect, roman symbol B), and one for interactive effects--the dependence of drug effect on subject. There appears to be little or no interactive effect, because the lines rise from left to right. In this analysis we assume no interactive effect. Note that not enough data were provided to estimate the interactive effect. We have 20 numbers, which allows us to estimate a mean difference between drugs, a mean difference for each subject, leaving $20-1-1-9 = 9$ degrees of freedom for the error term. If we estimate the interaction term there will be no degrees of freedom for the error term.

1. Construct model

The revised model is:

Write GLM: $T = \beta_o + \beta_D \cdot X_D + \beta_S \cdot X_S + \text{residual}$

S&R95 $T_{ijk} = \mu + \alpha_i + B_j + \epsilon_{ijk}$

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, $X_S = 0, 1...10$
These are labels (categories), not numbers on ratio scale.

Code model statement in statistical package according to the GLM

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

```
MTB> ANOVA 'T' = 'XS'  'XD'
MTB> GLM   'T' = 'XS'  'XD'
SUBC> fits c4;
SUBC> res c5.
```

The grand mean.

$\hat{\beta}_0 = 20^{-1} \Sigma T = 20^{-1} \cdot 30.8 = 1.54$ hours

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

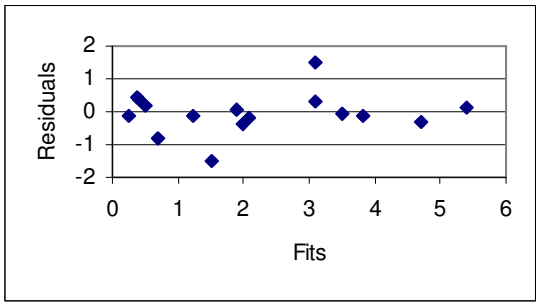
$\text{mean}(T_{D=A}) = 0.75$ hours
$\hat{\beta}_{D=A} = (0.75 - 1.54) = -0.79$ hours
$\text{mean}(T_{D=B}) = 2.33$ hours
$\hat{\beta}_{D=B} = (2.33 - 1.54) = +0.79$ hours
$\text{mean}(T_{S=1}) = 1.30$ hours
$\hat{\beta}_{S=1} = (1.30 - 1.54) = -0.24$ hours
$\text{mean}(T_{S=2}) = -0.40$ hours
$\hat{\beta}_{S=2} = (-0.40 - 1.54) = -1.94$ hours

etc.

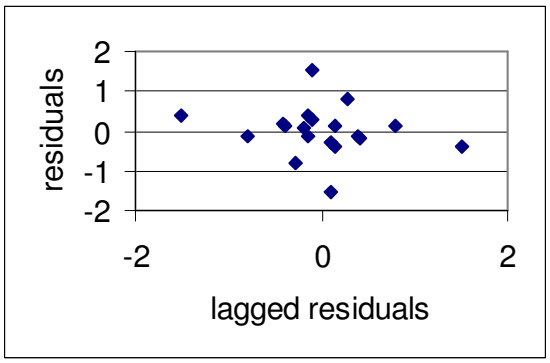
The table shows calculations in spreadsheet format.

		Drug	Subject		
Hrs		Effect	Effect	fits	res
T	βo	βD	βS	βo+βD+βS	
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.
Plot residuals versus fits.



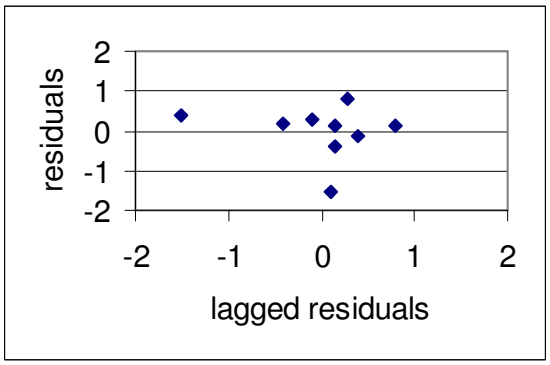
a. No line fitted in model,
so skip this evaluation.
b. No systematic change in residuals with increase in fitted values (*i.e.* no cones) so residual homogeneous, no need to revise error structure of model.
c. If n small, evaluate assumptions for computing p-values.
n = 48 so only large violations will distort p-values or confidence limits.



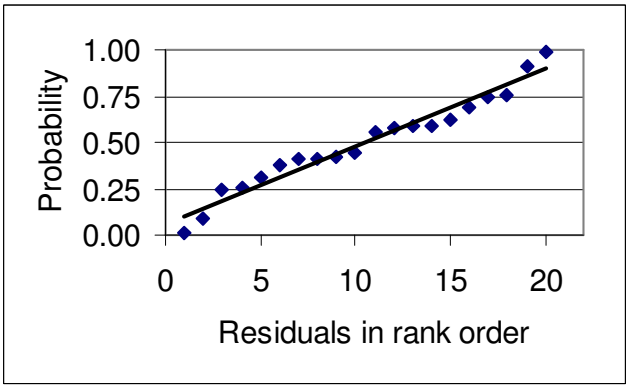
Homogeneous? Yes

Sum(res) = 0? Yes

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.



Normal ?
The residuals deviate only slightly from normal, as judged relative to trend line in the normal probability plot. The deviations are too small to conclude that the residuals are not normal.

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 drugs were set by experimental design. We will view drug as a fixed factor and hence infer only to these two drugs. The subjects were not selected. They were chosen haphazardly, or better yet, at random from a larger population. Based on the characteristics of the sample, we are prepared to infer to a population of subjects with similar characteristics. We would require a well executed clinical trial on non-institutionalized subjects before making claims about the efficacy of one drug relative to another in a larger population.

We can infer to a population of all possible measurements of hours of extra sleep, given the mode of collection. Beyond that, we can infer to a population of subjects with characteristics similar to those in the study.

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

Research question is binary: do drugs differ in effect, controlling for individual variation in response to the drugs. So hypothesis testing is appropriate (step 6).

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

We assume no interactive effects, i.e, we assume that the effect of one drug relative to another is consistent across subjects. Hypothesis testing will focus on the drug effect. We are not interested in whether subjects differ. The subject factor, be being present 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: Mean($T_{D=A}$) ≠Mean($T_{D=B}$) The population means differ

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

These hypotheses are equivalent to

H_A: $\beta_D \neq 0$

H_o: $\beta_D = 0$

Are there more specific hypotheses about parameters? No

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.

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

GLM	$T - \beta_o$	=	$\beta_S \cdot X_S$	+	$\beta_D \cdot X_D$	+	ϵ
Source	Total	=	Subject	+	Drug	+	Resid
df	20−1	=	10−1	+	2−1	+	19−9−1

Construct ANOVA table according to terms in model.
Take df from beneath each term and place in table.

Source	df	SS	MS	F	---->	p
Subject	9					
Drug	1					
Res	9					
Total	19					

7. ANOVA - Calculate then partition variance according to model.
Statistical packages produce SS for each term in the model, then report the results in the ANOVA table.

GLM	$T - \beta_o$	=	$\beta_S \cdot X_S$	+	$\beta_D \cdot X_D$	+	ϵ
Source	Total	=	Subject	+	Drug	+	Resid
SS	77.37	=	58.08	+	12.48	+	6.81

Source	df	SS	MS	F	---->	p
Subject	9	58.08				
Drug	1	12.48				
Res	9	6.81				
Total	19	77.37				

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

The first two and last two data equations are shown.

Mean, SS, and df are computed for each row (term) in the model.

SS and df are then moved to ANOVA table, where MS is calculated from SS/df.

The F-ratio is taken relative to drug. No F-ratio computed from random factor (subject).

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			

The p-value from the F-distribution is 0.0028.

8. Decide whether to recompute p-value.
No need, because deviations from normal errors were small. Even if the violations were large the decision would not be changed because the p-value is far from 5%.

9. Declare decision about terms.
Only one term, the fixed factor, is tested.
$p = 0.0028 < \alpha = 5\%$ reject H_o and accept $H_{A:}$ extra sleep depends on drug.

This two-way ANOVA is also known as a paired t-test. There is one random factor and the fixed factor has two categories. The paired t-test is also conducted by calculating the difference within each random category, computing the mean difference across categories, then testing whether the mean difference differs from zero. Here are the calculations.

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	$t = \frac{1.58}{\sqrt{(13.62/9)/10}}$	
	p	0.0028		

10. Report and interpret parameters of biological interest.
No parameters are reported for the subject term because it is a random factor and so the means are of no interest.

The confidence limits for each group overlap because the among subject variance is not controlled statistically.

	st.err	Lower CL	Upper CL
mean(T_A) = 0.75 hours	0.5657	−0.53 hours	2.03 hours
mean(T_A) = 2.33 hours	0.6332	+0.80 hours	3.06 hours

Next, the confidence limits for the average difference, controled for among subject variation.
Note the much smaller standard error after the among subject variation is removed.

	st.err	Lower CL	Upper CL
mean($T_B - T_A$) = 0.75 hours	0.3890	0.70 hours	2.46 hours

The confidence limits do not include zero; there is a significant difference between the drugs.