Model Based Statistics in Biology.
Part III. The General Linear Model.
Chapter 10.2 Two Sample t-test

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

Ch10.xls
Sleep data from Cushny and Peebles
Daphnia ages from Sokal and Rohlf (1995) Box 9.5

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.

Today:
Two-sample t-test as a special case of the GLM

Wrap-up
ANOVA a special case of the general linear model..
Explanatory variable on nominal scale.
Special case of one-way (single factor) ANOVA: two means. Called a t-test.

GLM, applied to ANOVA One way (single factor) ANOVA with two classes.
Example. Sleep data.

The example will be hours of extra sleep, relative to control,
for hyoscyamine (DrugA) and hyoscine-L (DrugB).
Data from
Cushny AR, Peebles AR (1905). The action of optical isomers. II. Hyoscines. J Physiology 32:501-510.

Used by W.S. Gossett in paper that introduced the t-test. (Student. 1908) and then by Fisher (1925) in the first text in statistics.

0.7	1.9
-1.6	0.8
-0.2	1.1
-1.2	0.1
-0.1	-0.1
3.4	4.4
3.7	5.5
0.8	1.6
0.0	4.6
2.0	3.4

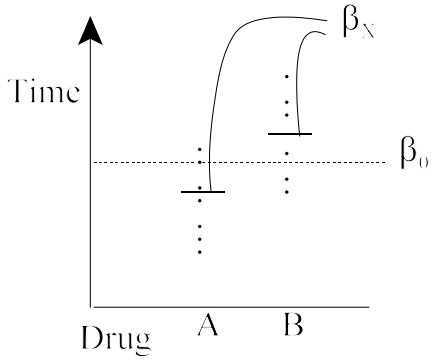
DrugA DrugB

Cushny.dat

1. Construct model

Verbal model:
Extra time slept depends on drug.

Graphical model



The verbal and graphical models help us distinguish response from explanatory variables.

First, define the quantity of interest.
This is the response variable.
Response variable: Hours of extra sleep *T* relative to control

Then, define the explanatory variables.
Explanatory variable: *Drug* (= A or B) (Drug is on nominal scale)

State type of measurement scale for response variable and explanatory variable(s).

	Variables name	Symbol	Units	Scale
Response variable	Hours of extra sleep	*T*	hours	ratio
Explanatory variables	Drug	*Drug*		nominal

Next, write the model using names of quantities.
Hours of extra sleep = f(Drug type).
"Hours of extra sleep depend on drug type"

Finally, write the model in more abstract form, which is what the computer will use to carry out the analysis.
Formal Model $T = \beta_o + \beta_{Drug}\, Drug + \epsilon$
β_0 is the grand mean (mean across both drugs).
β_{Drug} are the means for each group,
expressed as deviations from the grand mean
$\beta_0 + \beta_{Drug}$ = means of each group.

This model, which compares one mean to another, is called a t-test. This model is a single factor ANOVA with just 2 groups.

2. Execute analysis.

Place data in model format:

Column with response variable, extra sleep time T.

Column with explanatory variable, $Drug$ = 0 or 1

Code model statement in statistical package according to the GLM

$$T = \beta_o + \beta_{Drug}\, Drug + \epsilon$$

```
MTB> ANOVA 'T'' = 'Drug'
MTB> GLM 'T' = 'Drug'
```

If you are using a graphics interface to run the analysis, be sure to look at the code produced, so that you understand how the model you wrote translates into a model statement in your package.

Obtain fitted values and residuals.

Fitted values:

Fits = E[T] = $\hat{\beta}_o + \hat{\beta}_{Drug} \cdot Drug$

Residuals: Res = T – Fits

There are several ways to obtain residuals and fits.

a. GLM routines typically produce residuals and fits as output.

GLM: $T = \beta_o + \beta_{Drug}$ Drug + ϵ

```
MTB > GLM   'Time' =          'Drug';
SUBC> fits c3;
SUBC> res c4.
```

The model statement in Minitab looks much like the model statement in other packages.

2. Execute analysis.

b. Calculate fits and residuals from the parameters estimated directly. The parameter estimates are:

$\bar{T}_A = 0.75$
$\bar{T}_B = 2.33$

These are the fitted values. Compute residuals from difference between data and fitted values (means).

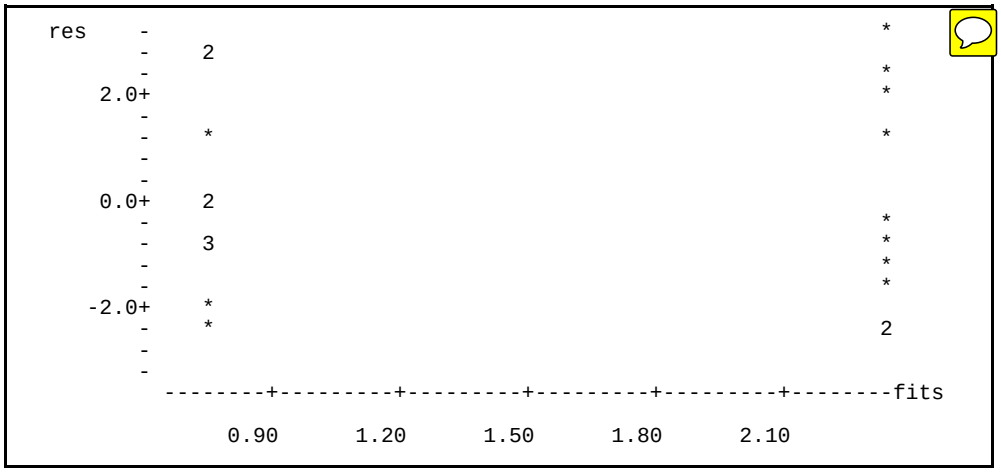
```
MTB > print 'T' 'Drug'  'fits' 'res'

 ROW      T   Drug   fits    res

   1    0.7      1   0.75  -0.05
   2   -1.6      1   0.75  -2.35
   3   -0.2      1   0.75  -0.95
   4   -1.2      1   0.75  -1.95
   5   -0.1      1   0.75  -0.85
   6    3.4      1   0.75   2.65
   7    3.7      1   0.75   2.95
   8    0.8      1   0.75   0.05
   9    0.0      1   0.75  -0.75
  10    2.0      1   0.75   1.25
  11    1.9      2   2.33  -0.43
  12    0.8      2   2.33  -1.53
  13    1.1      2   2.33  -1.23
  14    0.1      2   2.33  -2.23
  15   -0.1      2   2.33  -2.43
  16    4.4      2   2.33   2.07
  17    5.5      2   2.33   3.17
  18    1.6      2   2.33  -0.73
  19    4.6      2   2.33   2.27
  20    3.4      2   2.33   1.07
```

c. Calculate fits and residuals from the parameters estimated by GLM routine. The parameter estimates from the GLM routine are:

$\hat{\beta}_0 = 1.54$

$\hat{\beta}_{Drug} = \begin{matrix} +0.79 \\ -0.79 \end{matrix}$ $\quad \bar{T}_A = 1.54 - 0.79 = 0.75$ $\quad \bar{T}_B = 1.54 + 0.79 = 2.33$

Plot residuals vs fits

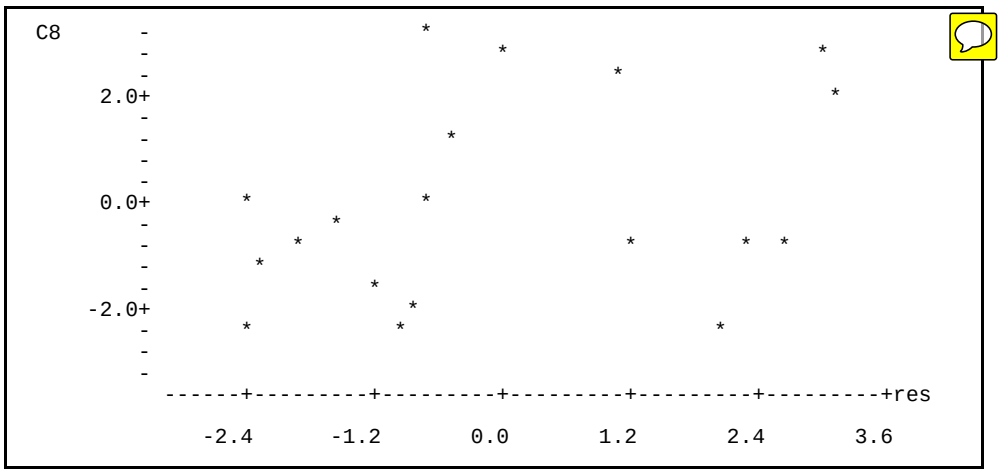


3. Evaluate the model.

a. No slopes (straight lines) used, so no need to check straight line assumption.
b. The residuals appear homogeneous, they form stacks of similar vertical dispersion, one stack for each group (each drug).
c. If n small, evaluate assumptions for using chisquare, t, or F distribution
First assumption: Homogeneous errors ? Yes.
Second assumption is met whenever parameters are estimated by least squares. Statistical packages all use least squares, so this assumption met when these packages are used.

3. Evaluate the model.

Third assumption. Independent errors?



independent ?

Yes

Fourth assumption. Normal errors

```
MTB > hist 'res';
SUBC> incr 1.
Histogram of res   N = 20
Midpoint   Count
   -2.00       5  *****
   -1.00       5  *****
    0.00       3  ***
    1.00       2  **
    2.00       2  **
    3.00       3  ***
```

Normal ? No

Residuals in these plots look homogeneous, independent, but not normal
It is interesting to note that the data used by Gossett (1908) to introduce the t-test did not meet the assumptions for this test.
The residuals departed noticeably from normal so we may need to recompute p-value from an empirical distribuion rather than the t-distribution introduced by Gossett.

4. State sample and population

The population in this case is all possible differences in sleep between two groups. This population is represented by the term $\hat{\beta}_{Drug}$ in the model. We could estimate this term by running the experiment repeatedly, them taking the average value of the differences between the two groups. From this point of view, the population is all possible measurements under the conditions specified in the measurement protocol for this experiment. The sample is considered applicable to any repeat of the experiment, conducted in the same way.

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

Yes. The question is whether one drug is better than the other. It is not at all clear whether the greater hours of sleep due to the one drug is more than just chance.

6. State H_A / H_o pair, with tolerance for Type I error.

There is one term in the model. Is this term significant ? (not due to chance).

The research hypothesis is that the drugs differ in effect. H_A: $\mu_A \neq \mu_B$
The null hypothesis is the drugs do not differ in effect. H_o: $\mu_A = \mu_B$

The symbol β_{Drug} has a single value, the difference between the two means.
The symbol β_{Drug}·Drug has two values,
zero and the difference between the two means.
The hypotheses listed above are equivalent to the following pair of hypotheses
If the means differ, then var(β_{Drug}·Drug) >0 H_A: var(β_{Drug}·Drug) > 0
If the means are the same,
then is no variance in β_{Drug}·Drug H_o: var(β_{Drug}·Drug) = 0
For this analysis, we have no additional hypotheses concerning parameters.
(For example, that one drug is better than another).

State test statistic	F-ratio
Distribution of test statistic	F-distribution
Tolerance for Type I error	5%

7. ANOVA - Partition df according to model.

```
ANOVA table headings on chalk board, upper right.
GLM just to the left.
Headings under model, then move to ANOVA table
```

GLM:	A	=	β_o	+	β_{Drug} Drug	+	ϵ
Source:	Total	=			Drug		residual

Source	df	SS	MS	F	---->	p
Drug						
Res						
Total						

Partition df according to model

GLM:	A	=	β_o	+	β_{Drug} Drug	+	ϵ
Source:	Total	=			Drug		residual
df	20	=	1	+	1	+	18

GLM:	A – β_o	=			β_{Drug} Drug	+	ϵ
Source:	Total	=			Drug		residual
df	20 – 1	=			1	+	18

Source	df	SS	MS	F	---->	p
Drug	1					
Res	?		? = 19 – 1 = 18			
Total	19					

7. ANOVA - Partition SS according to model.

Calculate SS_{tot} from Var(Time), the variance of response variable.

SS_{tot} = Var(Time) $\cdot$ df_{tot} = 4.072 $\cdot$ 19 = 77.37

Then partition SS_{tot} according to model. Model statement does this.

GLM:	A – β_o	=	β_{Drug} Drug	+	ϵ
Source:	Total	=	Drug		residual
df	20 – 1	=	1	+	18
SS	77.37	=	12.48	+	64.89

Source	df	SS	MS	F	----> p
Drug	1	12.48			
Res	18	64.89			
Total	19	77.37			

Table source, df, SS, MS, F-ratio

```
Moving from left to right,
  compute MS from SS and df in ANOVA table
  compute F from MS
```

Next, compute Mean Squares MS from SS and df

MS stands for the mean squared deviation. MS = SS / df

MS_{model} = MS_{Drug} = 12.482 add to table.

MS_{res} = MS_{error} = 3.605 add to table

MS_{tot} = Var(response) = Var(*T*)

not usually calculated or placed at bottom of MS column,

because $MS_{model} + MS_{res} \neq MS_{tot}$

Compute test statistic $F = (MS_{Drug})/(MS_{res}) = (SS_{Drug})/(SS_{total} - SS_{Drug})$

The F-ratio can be thought of as the signal to noise ratio.

How strong is the signal, relative to the noise (error) ?

Add F to table.

Typically F is the ratio of the explained variance (due to the entire model, or due to a factor in the model) to the unexplained variance.

E.g., $F = MS_{model} / MS_{res}$

This can be calculated by hand, if necessary, using MS or SS from computer package.

Forming the correct F-ratio can require considerable skill and experience, especially with complex designs.

Computer packages sometimes produce incorrect F-ratios. It is a good idea to check with a statistician, if in doubt.

Source	df	SS	MS	F	----> p
Drug	1	12.48	12.482	3.46	
Res	18	64.89	3.605		
Total	19	77.37			

The completed table represents a sequence of computations from left to right. It results in an F-ratio, which will be small if the H_o is true, and will be large if the H_A is true.

7. ANOVA Table source, df, SS, MS, F-ratio, and p-value.

Here are the same computations in a spreadsheet.

Time	Drug	Fits	Residuals	
0.7	0	0.75	-0.05	
-1.6	0	0.75	-2.35	
-0.2	0	0.75	-0.95	
-1.2	0	0.75	-1.95	
-0.1	0	0.75	-0.85	
3.4	0	0.75	2.65	
3.7	0	0.75	2.95	
0.8	0	0.75	0.05	
0	0	0.75	-0.75	
2	0	0.75	1.25	
1.9	1	2.33	-0.43	
0.8	1	2.33	-1.53	
1.1	1	2.33	-1.23	
0.1	1	2.33	-2.23	
-0.1	1	2.33	-2.43	
4.4	1	2.33	2.07	
5.5	1	2.33	3.17	
1.6	1	2.33	-0.73	
4.6	1	2.33	2.27	
3.4	1	2.33	1.07	
4.072		0.657	3.415	variance
77.368	=	12.482	64.886	SS= 19*variance
19.000		1.000	18.000	df
		12.482	3.605	MS

SS_{tot} = Var(T)* df_{tot}
SS_{fits} = Var(fits) * df_{tot}
SS_{res} = Var(res) * df_{tot}
The same computations can be carried out in any package

Calculate p-value for terms in model.

```
MTB > cdf 3.46;
SUBC> f 1 18.
  0.921    3.46
```

Hence p = 1 – .921 = 0.079

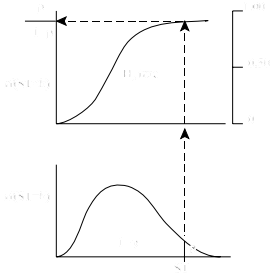


Fig L10a

8. Recompute p-value if necessary.

When assumptions not met, recompute if: n small (Yes, n = 19)
p near α (Yes, p =0.079)

p-value near α, so decision might change if p-value recomputed by randomization.

Colquhoun (1971) carried out a randomization test, using
12000 of the 184756 possible permutations of the data into 2 groups.
The p-value was p = 0.0813 (976/12000)
Close to p-value from t-test, leaving decision unchanged.

The p-value changed by a factor of 0.0813 / 0.079 = 1.03 (hardly at all)

Usually it is not feasible to construct frequency distribution from all permutations. Instead, we sample from the list of all permutations by sampling at random from the data, and computing the F-ratio repeatedly to construct the frequency distribution of F when the null hypothesis is true.

9. Declare and report decision about model terms (compare p to α).

(*I.e.* Compare the observed statistic to population of such statistics)
0.0813 =p > α = 0.05 so accept H_o: Var(β_{Drug}) = 0
Equivalent to accept H_o: $\beta_{Drug} \neq 0$
Report decision:
There is no significant difference in extra time slept, for the two drugs
$F_{1,18}$ = 3.46 p = 0.081 (randomized)

When we accept the null hypothesis, we need to consider Type II error, that of accepting a null hypothesis that is not true.

We ask: What difference could have been detected, given the variance and the sample size? To answer this, we take the observed difference between the means (ΔT = 2.33 – 0.75 = 1.58 hours), then increase this difference until the p-value becomes significant. We start with guess : we increase difference by adding 0.5 hours to each value in the group with the larger mean. This increases the mean to 2.83 hours. It increases the difference to ΔT = 2.08 hours. Then we run the GLM routine to obtain the p-value.

ΔT = 2.33 – 0.75 = 1.58	F = 3.46	p = 0.079	p-value from F-distribution
ΔT = 2.83 – 0.75 = 2.08	F = 6.00	p = 0.025	too high, try 0.2 increase
ΔT = 2.53 – 0.75 = 1.78	F = 4.39	p = 0.05	

The minimum detectable difference was 1.78 hours, which is higher than the observed difference by 1.78/1.58 = 1.127. With this sampling effort and variance we could have detected a difference of 1.78 hours. A true difference of 1.77 hours of extra sleep would go undetected by this experiment. A better experiment is needed, one that has a chance of detecting a smaller a difference. One way to improve the experiment is to increased the number of trials, which will reduce the error variance.

9. Declare and report decision about model terms (compare p to α).

If we are planning another experiment it is informative to compute the sample size needed to detect a difference, given the variance and contrast between means. To do this we increase sample size until the F-ratio becomes significant. Because p already close to α we start with a small increase of 10, from n = 20 to n = 30.

Source	df	SS	MS	F	---->	p
Drug	1	12.48	12.482	4.52		0.0425
Res	28	77.37	2.763			
Total	29	64.89				

Try a slightly smaller increase, of 8 (4 per group), from n = 20 to n = 28

Source	df	SS	MS	F	---->	p
Drug	1	12.48	12.482	4.19		0.051
Res	26	77.37	2.976			
Total	27	64.89				

Assuming the same variance and same difference in means, a sample size of 15 per group (n = 30) was needed to detect a observed difference. This is a feasible increase

10. Report and interpret parameters of biological interest.

Parameters are not of interest because we conclude there is no difference.

```
MTB > describe 'hrs'; by 'drug'
              drug       N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
 hrs             1      10    0.750    0.350    0.675    1.789    0.566
                 2      10    2.330    1.750    2.237    2.002    0.633
```

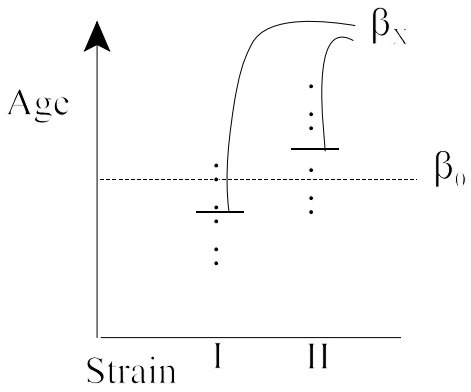
With this sampling effort and variability, we could have detected a difference of 1.78 hours in time of sleep, only about 13% higher than the observed by a factor of 1.78/1.58 = 1.127. Any effect smaller than 1.78 hours would go undetected with this sampling effort and variability. The study needs to be repeated to be conclusive.

Example. Daphnia ages
Data from Box 9.5 p 220 Sokal and Rohlf 1995
Does time to maturity differ in two genetic crosses in the water flea *Daphnia* ?

1. Construct model

Verbal model: age depends on strain.

Graphical model



L15aF2

Response variable

A = age (in days) at beginning of reproduction in *Daphnia longispina* in two genetic crosses I and II (ratio type of scale)

Explanatory variable.

X = I or II (nominal scale)

n = 14 observations, 7 in each of groups I and II

Formal Model $A = \beta_o + \beta_X X + \epsilon$

β_0 is the grand mean

β_X are the means for each group, expressed as deviations from the grand mean

$\beta_0 + \beta_X$ = means of each group.

2. Execute analysis. Place data in model format:

Column with response variable, Age A.

Column with explanatory variable, Strain = 0 or 1

Code model statement in statistical package according to the GLM

$$A = \alpha + \beta_{Strain} \cdot Strain + \epsilon$$

```
MTB> ANOVA 'A'' = 'Strain'
MTB> GLM 'A' = 'Strain'
```

Fits and residuals via any of several methods:

-model statement output

-direct calculation of parameters (two means)

-parameters reported by GLM routine

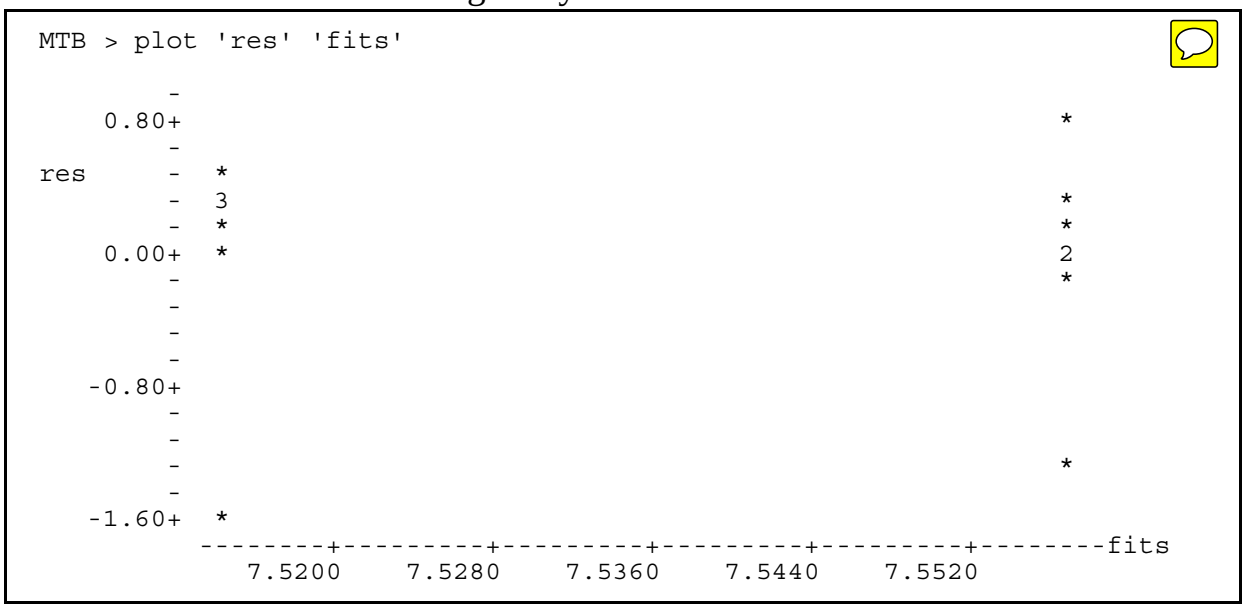
$\hat{\beta}_0 =$ 7.5357

$$\hat{\beta}_X = \begin{matrix} +\ 0.0214 \\ -\ 0.0214 \end{matrix} \qquad \begin{matrix} \bar{A}_I = 7.5357 + 0.0214 = 7.5571 \text{ days} \\ \bar{A}_{II} = 7.5357 - 0.0214 = 7.5143 \text{ days} \end{matrix}$$

3. Evaluate model

a. No slopes (straight lines) used, so no need to check for bowls/arches.

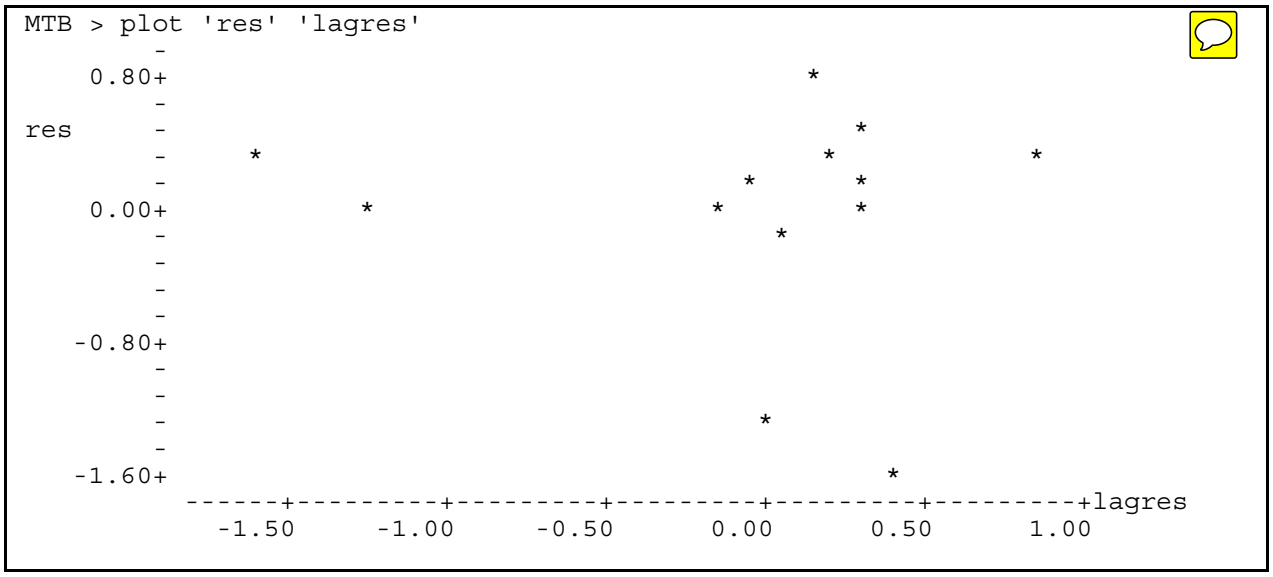
b. Examine residuals for homogeneity and outliers.



The two stacks of residuals in this plots are of similar spread, so we conclude the residuals are homogeneous.

3. Evaluate model

c. use residuals to decide if t-distribution can be used.



independent ? Yes

Normal ? No

```
MTB > hist 'res'

Histogram of res   N = 14

Midpoint   Count
    -1.6       1  *
    -1.2       1  *
    -0.8       0
    -0.4       0
     0.0       5  *****
     0.4       6  ******
     0.8       1  *
```

Distributional assumptions not met (residuals not normal) so may need to recompute p-value.

4. State population and whether sample is representative

Population not stated in text example. We assume it is all possible measurements, if this experiment were carried out repeatedly. I.e., hypothetical population of all possible measurements under the conditions specified in the measurement protocol for this experiment. There is an implicit assumption that results apply to biological populations, all *Daphnia* belonging to these two strains.

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

Hypothesis testing is warranted because it is evident from the data that the distributions around the two means overlap. The difference might be due to chance.

6. State H_o/H_A pair

There is one term in the model. The hypothesis pair is for this term.

The mean age differs in the two strains (I and II). H_A: $\mu_I \neq \mu_{II}$

The mean age does not differ in the two strains. H_o: $\mu_I = \mu_{II}$

This is equivalent to a test of whether there is variance in the two means.

There is variance in the means of the groups. H_A: var(β_{Strain}*$Strain$) $\neq 0$

There is no variance in the means of the groups. H_o: var(β_{Strain}*$Strain$) $= 0$

Yet another way to state the same H_o/H_A pair is with the t-statistic.

The mean age differs in the two strains. H_A: $t \neq 0$

The mean age does not differ in the two strains. H_o: $t = 0$

The formula for the t-statistic (equal sample size) is

$$t = \frac{\left(\bar{X}_I - \bar{X}_{II}\right) - \left(\mu_I - \mu_{II}\right)}{\sqrt{\frac{1}{n}\left(s_I^2 + s_{II}^2\right)}}$$

The Type I error is set at the conventional 5% level.

7. ANOVA - Partition df and variance according to model.

```
ANOVA table headings on chalk board, upper right.
GLM just to the left.
Headings under model, then move to ANOVA table
```

Calculate $df_{total} = n - 1 = 14 - 1 = 13$ Partition df according to model

Calculate SS_{tot} from Var(A), the variance of response variable.

SS_{tot} = Var(A) $\cdot$ df_{tot} = 0.42247$\cdot$ 13 = 5.49214

Use statistical package to partition SStot according to model

GLM:	A – β_o	=	β_X X	+	ϵ
Source:	Total	=	Strain	+	residual
df	13	=	12	+	1
MTB > ANOVA	'age'	=	'strain'		
MTB > GLM	'age'	=	'strain'		
	5.492	=	0.00641	+	5.4857

Move partitioned df and SS to ANOVA table

Source	df	SS	MS	F
strain	1	0.00641		
error	12	5.48571		
total	13	5.49212		

```
Moving from left to right,
  compute MS from SS and df in ANOVA table
  compute F from MS
```

7. ANOVA - Table SS, MS, F-ratio.

Compute MS from SS and df

Compute test statistic $F = (SS_{gr})/(SS_{within}) = (SS_{gr})/(SS_{total} - SS_{gr})$

Source	df	SS	MS	F
strain	1	0.00641	0.00641	0.014
error	12	5.48571	0.45714	

The F-ratio, by definition, is t^2

$$t = \frac{(\bar{X}_I - \bar{X}_{II}) - (\mu_I - \mu_{II})}{\sqrt{\frac{1}{n}(S_I^2 + S_{II}^2)}} \qquad t = \frac{(7.5571 - 7.5143) - (0 - 0)}{\sqrt{\frac{1}{7}(0.50476 + 0.40952)}} = \frac{0.4286}{\sqrt{\frac{0.9143}{7}}} = \frac{0.4286}{0.3614} = 0.1186$$

$t^2 = 0.1186^2 = 0.014$

7. ANOVA - Calculate p-value for terms in model.

```
MTB > cdf .014;
SUBC> f 1 12.
  0.014    .0922
```

Hence p = 1 – .0922 = .9078

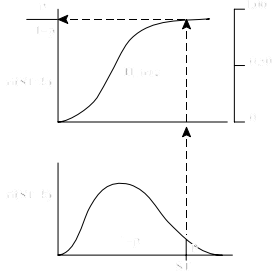


Fig L10a

8. Decide whether to recompute p-value by randomization

Do we need to recompute the p-value ?

No, because the p-value is far from the criterion level of α. The p-value won't change enough to alter the decision to accept the null hypothesis.

```
p-value was computed by randomization in lab4 for this example,
to allow comparison with p-value from F-distribution.
Compile p-value from all randomizations done by one class.
```

The assumption of normal residuals has the least affect on the p-value calculated from the F-distribution. In general, non-homogeneous residuals are a more serious problem.

9. Make decision by comparing p-value to criterion α

$0.91 = p > \alpha = 0.05$

so accept H_o: There is no significant difference in mean age.

There is no significant difference in mean age between the two strains of *Daphnia*.

($F_{1,12} = 0.014$ $p = 0.91$)

Report statistic and sample size. Report p, not alpha.

The observed difference in age can easily have been observed by chance alone.

9. Make decision by comparing p-value to criterion α

When we accept the null hypothesis we need to consider Type II error, that of accepting the null when in fact the null is not true. In general we do not compute Type II error because it gives us no information beyond what we already have (the p-values). However, it is of interest to consider how much a difference we could have detected, given our data.

What is the minimum difference that could have been detected, given the variance and sample size?

To do this we keep increasing the difference between two groups until the difference is significant. In practice we add an offset to one group, compute the t-statistic and p-value, increase the offset, compute the t-statistic and p-value again, and continue until the p-value falls below the significance level (5%).

This computation can be done in a spreadsheet.

Age	Strain	Fits	Residuals	Strain 0	Offset
8.002	0	8.3163	-0.31429	7.2	0.802
7.902	0	8.3163	-0.41429	7.1	0.802
9.902	0	8.3163	1.585714	9.1	0.802
8.002	0	8.3163	-0.31429	7.2	0.802
8.102	0	8.3163	-0.21429	7.3	0.802
8.002	0	8.3163	-0.31429	7.2	0.802
8.302	0	8.3163	-0.01429	7.5	0.802
8.8	1	7.5571	1.242857		
7.5	1	7.5571	-0.05714		
7.7	1	7.5571	0.142857		
7.6	1	7.5571	0.042857		
7.4	1	7.5571	-0.15714	Strain 0	8.3163
6.7	1	7.5571	-0.85714	Strain 1	7.5571
7.2	1	7.5571	-0.35714	Diff	0.7591
0.577		0.155	0.422	variance	
7.503	=	2.017	5.486	SS= 13*variance	
13.000		1.000	12.000	df	
		2.017	0.457	MS	
			4.412281	F	
			0.050037	p	

The two strains would have to differ by 0.76 days to be statistically significant.
I.e. the strains would have to differ by (0.7591/7.5571) = 10% to be significant.
The analysis was capable of detecting a 10% difference in age.
The absence of a significant difference cannot be attributed to a poorly executed study.

10. Report and interpret parameters of biological interest.

We conclude there is no difference in age between the two strains.
The difference between the two strains will not be of biological interest.

$\hat{\beta}_0$ = 7.5357 standard error = 0.174, n = 14. $\bar{A}_I$ = 7.5571 days

$\bar{A}_{II}$ = 7.5143 days

The two means differ by only 6 parts in a 1000 (7.5571 – 7.5143)/7.5357 = 0.006
The parameter of biological interest is the average time to maturity, regardless of strain, which is $\hat{\beta}_0$ = 7.5357days sterr = 0.174 days n = 14