

## Model Based Statistics in Biology.

### Part III. The General Linear Model.

#### Chapter 10.1 Single 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_1a.xls	Extra sleep(hyoscamine)
Ch10_1b.xls	Extra sleep (work in class)
Ch10_2.xls	Daphnia ages SRBx9_5.out

Ch10_1b.xls
Example to be worked in class, from data collected in class, using CI

#### **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 example 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:
--------

Single sample t-test as a special case of the GLM
---

Move single sample t-test to goodness of fit chapter. Goodness of fit to chance.
---

#### **Wrap-up**

The single sample *t*-test compares an observed mean to a population mean of no difference.

The sample consists of random units (10 individuals in this case) for which a difference can be calculated.

### Single Sample t-test.

The *t*-test was developed over a century ago by W.S. Gossett, who worked for the Guinness brewery in Dublin, Ireland. Gossett was concerned with the problem of small sample sizes in choosing the best yielding varieties of barley. Gossett worked out the formula for the standard error of the mean for small sample sizes, Gossett published the derivation under a pseudonym because his employers were concerned with the publication of trade secrets.

Student. 1908. The probable error of a mean. *Biometrika* 6: 1-25.

Instead of barley data Gossett used physiological data on drug efficacy.

Cushny AR, Peebles AR (1905). The action of optical isomers.

II. Hyoscines. *J Physiology* 32:501-510.

0.6	1.3	0.7
3.0	1.4	-1.6
4.7	4.5	-0.2
5.5	4.3	-1.2
6.2	6.1	-0.1
3.2	6.6	3.4
2.5	6.2	3.7
2.8	3.6	0.8
1.1	1.1	0.0
2.9	4.9	2.0
Control	DrugA	Diff
Cushny.dat		

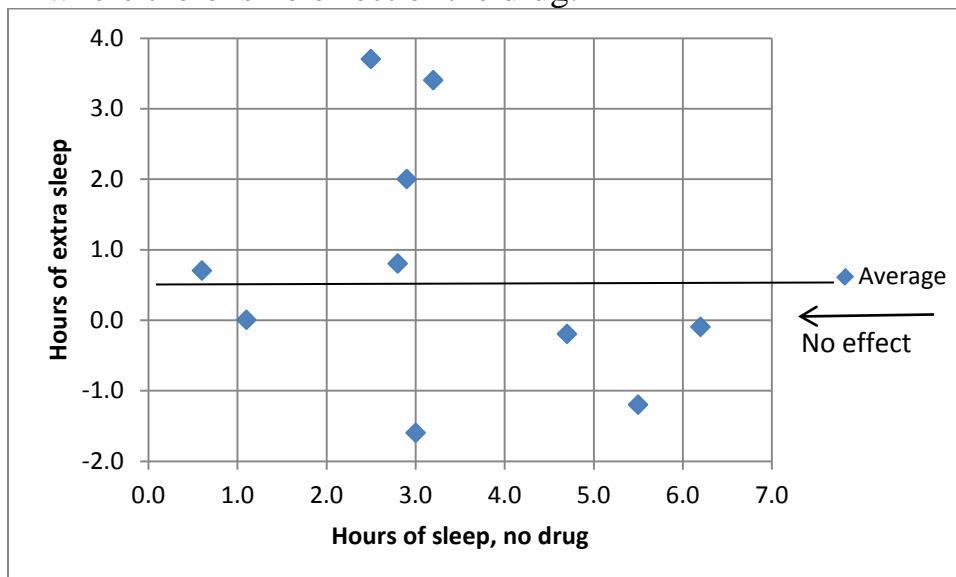
The research question was whether Drug A (L-hyoscyamine) results in extra sleep relative to no drug, in 10 subjects. Through a labelling error, Gossett reported the L-hyoscyamine data as dextro-hyoscyamine.

R.A. Fisher, an applied statistician at the Rothamsted agricultural research station, appreciated the importance of Gossett's derivation when cost of replication is high. Fisher used the data from Gossett's publication in the first text in statistical methods (Fisher1925 Table 27) .

### 1. Construct model

Verbal model: Drug A increases time slept.

Graphical model: Compare the average of the sample to the average in a population where there is no effect of the drug.



# 1. Construct model

Formal model.

Verbal and graphical models guide us in constructing the formal (symbolic) model.

To start, we define the quantity of interest.

$$\text{Hours of extra sleep for Drug A } T = T_{drugA} - T_{control}$$

This is the response variable

Then we define the explanatory variables. The explanatory variable here is Drug A versus a control, no drug. Because we have measurements of extra sleep both with and without the drug in the same individual, we can reduce the effect of the variable to a single number, the difference in hours of extra sleep.

$$T_{drugA} - T_{control} = \beta_o$$

$\beta_o$  the value of the difference in a population where there is no drug effect.

$\hat{\beta}_o$  is an estimate of  $\beta_o$  from a sample from such a population.

$\hat{\beta}_o$  will differ from  $\beta_o$  due to uncontrolled sources of error.

State type of measurement scale for response variable and explanatory variable.

<u>Variables</u>	<u>Name</u>	<u>Symbol</u>	<u>Units</u>	<u>Scale</u>
Response variable	Extra sleep measured	$T$	hours	ratio
Explanatory constant	Extra sleep in population	$\beta_o$	hours	ratio

Distinguishing the response from the explanatory variable is the first step in deciding on the statistical analysis to use.

Response variable

if ratio

then GLM

Explanatory variable

if single sample (class)

single population t-test

if two classes

t-test

if two or more classes (nominal)

then ANOVA

if ordinal

then ANOVA

if ratio

then regression

if ratio and nominal

then ANCOVA

Having distinguished the response from explanatory variable, we can write the model in more abstract form

$$\text{Formal Model } T = \beta_o + \epsilon$$

$\beta_o$  is the mean value of extra sleep, if drug A has no effect

This model, which compares an observed mean to a fixed value  $\beta_o$ , is called a one sample t-test. We can think of it as a goodness of fit test to a constant, zero extra sleep.

## 2. Execute analysis.

We will use data equations to illustrate the calculation of the single sample t-test  
Place data in model format: Response variable  $T$  in a column

Obtain fitted values.

Fitted value = average( $T$ ) = 0.75

Obtain residual values

Residuals:  $\text{Res} = T - \text{Fits}$

$$34.43 = SS_{\text{total}} = \sum T^2$$

$$28.81 = SS_{\text{residual}} = \sum (T - \bar{T})^2$$

5.63 = The improvement in fit

hrs treatment	hrs control	Trt-Ctl T	fits	res	res <sup>2</sup>
1.3	0.6	0.7	0.75	-0.05	0.002
1.4	3.0	-1.6	0.75	-2.35	5.523
4.5	4.7	-0.2	0.75	-0.95	0.903
4.3	5.5	-1.2	0.75	-1.95	3.803
6.1	6.2	-0.1	0.75	-0.85	0.723
6.6	3.2	3.4	0.75	2.65	7.023
6.2	2.5	3.7	0.75	2.95	8.703
3.6	2.8	0.8	0.75	0.05	0.003
1.1	1.1	0.0	0.75	-0.75	0.563
4.9	2.9	2.0	0.75	1.25	1.563
Sum				0.00	28.81

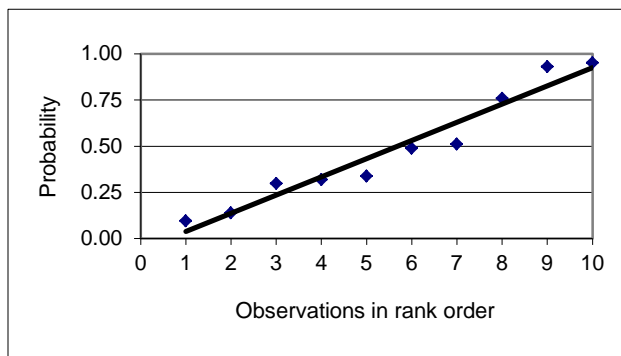
## 3. Evaluate the model.

1. No slopes (straight lines) in this model, so no need to check straight line assumption.
2. If n small, evaluate assumptions for normal probability model.

2a. Homogeneous errors ?

Not applicable because there is only one sample at fitted value of 0.75 hrs.

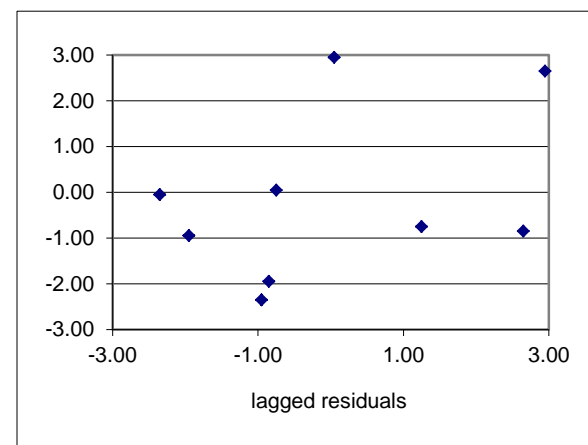
2b. Normal errors ?



Residuals depart little from straight line in the probability plot

2c. Independent errors? Yes, if we take the values in the order presented.

We see no trend in plot of residuals versus residuals at lag 1 (neighboring residual).



2b. Errors sum to zero? This assumption is met when parameters are estimated by least  
Statistical packages use estimation routines that result in errors that meet this assumption.

**4. Partition df and SS according to model.**

There are 10 observations, hence 10 df. Of these, one is lost upon estimating the value of the average difference  $\beta_o$

$$5.625 = SS_{\text{model}} = 10 \cdot (0 - 0.75)^2$$

$$28.81 = SS_{\text{residual}} = \sum (T - \bar{T})^2 = 28.81$$

$$34.44 = SS_{\text{total}}$$

GLM:	T	=	$\beta_o$	+	$\epsilon$
df	10	=	1	+	9
SS	34.435	=	5.625	+	28.81

**4. Calculate likelihood ratio.**

$$LR = (28.81/34.435)^{(-10/2)} = 2.4$$

The observed difference (0.75 hours) is only 2.4 times more likely than 0 extra hours. There is little evidential support for the alternative model.

**4. State the full (null) and reduced (alternative) model pair**

The research hypothesis is that the drugs differ in effect.

The average difference  $\beta_o$  differs from 0.  $H_A: \beta_o \neq 0$

We consider both tails because it is possible the drug reduces sleep, leading to negative values.

The null hypothesis is that the drug has no effect.  $H_o : \beta_o = 0$

**5. State sample and population**

The population in this case is all possible differences in time slept with and without the drug, given the experimental protocol, including type of drug and manner of administration of the drug. This is a hypothetical, not an enumerable population. We could estimate the average difference ( $\beta_o \hat{}$ ) by running the experiment repeatedly, then taking the average value of these differences. 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. We want to infer to a population of repeated use of the drug. We wish to avoid a false positive, resulting in prescribing an ineffective drug.

**6. State test statistic, its distribution (t or F), and tolerance of Type I error..**

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

**7. ANOVA - Calculate df according to model.**

Calculate mean squares MS from SS and df

$$\begin{aligned} MS_{res} &= 3.201 & SS / df &= 28.81 / 9 \\ MSD_{Drug} &= 5.625 & 10(\hat{\beta}_0 - 0)^2 &= 10(0.75)^2 \end{aligned}$$

Calculate test statistic

$$\begin{aligned} F &= 1.76 & (MS_{Drug}) / (MS_{res}) &= 5.625 / 3.201 \\ t &= 1.33 & t^2 &= F \end{aligned}$$

The F-ratio will be large if the average difference between drugs is large.  
The F-ratio will be large if  $MS_{residual}$  is small

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

```
MTB > cdf 1.76;
SUBC> f 1 9.
0.782 1.76
```

Hence  $p = 1 - 0.782 = 0.218$

```
MTB > cdf 1.3257;
SUBC> t 9.
0.891 1.3257
```

$p = 2(1 - 0.891) = 2(0.109) = 0.218$

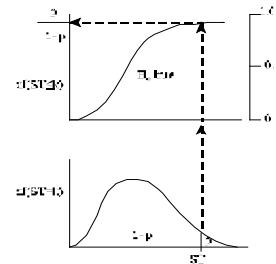


Fig L10a

**8. Recompute p-value if necessary.**

Assumptions met so no need to recompute Type I error.

**9. Declare and report decision about model terms (compare p to  $\alpha$ ).**

That is, compare the observed statistic to a population of such statistics.

$0.218 = p > \alpha = 0.05$

Report decision:

We cannot reject the null hypothesis of no effect for drug A

$F_{1,9} = 1.76 \quad p = 0.218$

**10. Report and interpret parameters of biological interest.**

With this sampling effort and variability, we could have detected an increase of 1.29 hours in time of sleep. Any effect less than 1.29 hours would go undetected with this sampling effort and variability. The study needs to be repeated to be conclusive.

**Extension. What about Type II error?**

When we fail to reject the null hypothesis, we need to consider Type II error, that of failing to reject a false null hypothesis. Computing statistical power (1- Type II error) when the null hypothesis cannot be rejected provides no additional information. Statistical power is inversely related to the  $p$ -value: a test with a  $p$ -value of 0.10 has more power than a test with a  $p$ -value with 0.30. See: Hoenig, J.M. and D. M. Heisey. 2001 The Abuse of Power: The pervasive fallacy of power calculations for data analysis. *The American Statistician* 55: 19-24.

### Extension. What about Type II error?

If we are going to calculate what went wrong, we can calculate something more useful, a minimum detectable difference. We ask: What difference could have been detected, given the variance and the sample size? To answer this, we set up a spreadsheet that calculates  $p$ -values from the ANOVA table. We then increase the difference between treated and control by adding a constant to the 10 observed values of  $T_{\text{Drug}}$  until the  $p$ -value becomes significant. Here is an abbreviated table of results, starting with an initial guess of 1 hour instead of 0.75 hours.

$T = 0.75$ hours	$F = 1.76$	$p = 0.218$	
$T = 1.00$ hours	$F = 3.12$	$p = 0.111$	too low, try 1.25 hours
$T = 1.25$ hours	$F = 4.88$	$p = 0.054$	close, try 1.28 hours
$T = 1.28$ hours	$F = 5.12$	$p = 0.050$	1.29 hours significant at 5%

The minimum detectable difference was 1.28 hours, given the sampling effort and variance among subjects. An increase in hours of sleep less than 1.29 hours may exist, but went undetected at 5% criterion. This example reminds us that we don't "accept the null" when we can't reject it.

Why is the minimum detectable difference useful? We can compare it to our knowledge of hours slept per day and consider whether an extra hour or so is biologically reasonable. Someone getting 5 hours per day might say yes. So another experiment, with more subjects, should be considered before concluding there is no effect.

In considering another experiment an even more informative estimate is the sample size needed to detect a difference, given the variance and contrast between means. To do this we keep the observed mean at 0.75 hours and increase the sample size until the  $F$ -ratio becomes significant. We start with an increase from  $n = 10$  to  $n = 20$ .

$n = 10$	$F = 1.76$	$p = 0.218$	
$n = 20$	$F = 3.71$	$p = 0.0692$	close, try $n = 24$
$n = 24$	$F = 4.49$	$p = 0.0451$	close, but possibly too high, try $n = 23$
$n = 23$	$F = 4.30$	$p = 0.0501$	

Assuming the same variance and difference in means, a sample size of 24 was needed to meet the 5% criterion. The increase in sample size from 10 to 24 is certainly feasible.

So can we just increase the sample size by 14 and add the new data to what we already have? We can, but we cannot use our 5% criterion for null hypothesis testing on the augmented sample. Why not? For the same reasons that we protect ourselves from too many "significant"  $p$ -values by using Bonferroni corrections. We are doing multiple testing on the initial sample size of 10.