1. DESCRIBING DATA

Data (plural; singular = datum) basically come in two varieties: Quantitative data and Categorical data. Quantitative data may be subdivided into Continuous data (e.g. diameter of a neurone); and Discrete data (e.g. number of neurones in a ganglion). Categorical data are sub-divided into Ordinal data, that is ordered in categories of an incremental nature (e.g. semi-objective assessments of the strength of a reflex response on a 1-5 scale); and Nominal data where the numbers in a category do not lie along an incremental scale (e.g. males/females; blood type).

MEAN, STANDARD DEVIATION AND STANDARD ERROR OF THE MEAN

\[
\bar{x} = \frac{\sum x_i}{N}
\]

\[
\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}}
\]

These equations are the bread-and-butter of biomedical measurement of continuous variables: that is, those like height, weight, blood pressure etc that can (theoretically!) assume any real value between ± \( \infty \). The mean is the arithmetic average (sum of the measured values, \( x_i \), divided by the number of measurements, upper equation). The standard deviation, \( \sigma \), embraces part of the “normal” variation in the population described by a Gaussian equation.

\[
f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]
When height, weight or a host of other physiological variables are measured and graphed in the form of a frequency histogram, the result is often a ‘bell-shaped’ distribution resembling a graphical plot of the Gaussian equation. The key factors are the mean value, \( \mu \), and the constant \( \sigma \), which is called the “standard deviation”. The mean is the arithmetic average; and the standard deviation embraces part of the “normal” variation in the population on either side of the mean. Thus, the greater the normal variation in the population, the greater is the standard deviation. If the histogram of a population variable is adequately fit by a Gaussian distribution then about 68% of the measurements will be found lying within one standard deviation on either side of the mean; about 95% of values will lie within two standard deviations (\( \pm 2\sigma \)); and about 99% will lie within \( \pm 3\sigma \).

The standard deviation can also be roughly estimated from a sample (\( N \)) of measurements (\( x \)) using the equation:

\[
\sigma = \sqrt{[(\Sigma x^2/N) - \mu^2]}
\]

(This function is relatively easy to calculate from a set of data. If the number of measurements \( N \) is reasonably large - say about 30 – the estimate of \( s \) using this equation will be close enough to that given by the correct, full equation: within less than 2% of the true \( s \)).
FAQ: What is the difference between mean, median and mode?

- Mean is the value obtained by dividing the data values obtained by the number of samples (a simple "average").

- Median is the data value marking the mid-way point; that is, there are equal numbers of data points with values above and below the median.

- Mode or modal value, refers to the most common value, obtained by plotting a frequency histogram of the data.

It is quite common in biology or medicine, when measurements are frequently made by necessity using small samples from a much larger population, to assume that the distribution of values obtained can be fit by a Gaussian equation. However, this equation also implies that none of the estimates from small samples is likely to give a "correct" estimate of the "true" mean value. Another sample from the same population, and another and another, will usually all provide different mean values. In other words, any estimate of the true mean based on means of several samples will also be in error. The mean values calculated in this way will also have a Gaussian distribution and hence the "mean-of-means" will have a standard deviation. The standard deviation calculated for a set of mean values is referred to as the Standard Error of the Mean (or S.E.M.). The S.E.M. will usually be smaller than the standard deviation measured in any one sample from the population. Provided the number of measurements in any one sample is reasonably large, however, the S.E.M. can be estimated by:

$$S.E.M. = \frac{\sigma}{\sqrt{N}}$$

FAQ: What is the difference between variance and coefficient of variation?

Variance is the square of the standard deviation ($\sigma^2$). It scales linearly with the mean for a normal distribution: that is, other things being equal, doubling the mean will double the variance. Coefficient of variation is a normalised measure of the standard deviation: that is, standard deviation divided by the mean (C.V. = $\frac{\sigma}{X}$)

FAQ: What are ‘confidence limits’ (or confidence intervals)?

A measure of the degree of certainty about the value of a parameter. For example, for a large set of measurements described by a normal distribution, we can be certain with 95% confidence that the true mean lies between 1.96 times the standard deviation on either side of the mean calculated from the sample of data. When the samples are small however, the multiple of the number of standard deviations (or SEM's) required to give us 95% confidence of the value of the true mean increases.

FAQ: What is a "z-score"?

A measure of where the value of a parameter lies, in relation to the 95% confidence limits. For example a z-score of 1.96 means that the parameter lies just on the boundary of the 95% confidence limit for a normal distribution.
Caveats and Consequences: Lies, Damned Lies and Statistics...
A common mistake is to calculate the standard deviation of the means but then to do another calculation (using a spreadsheet function or an equation like the above) to calculate the “standard error”. This may yield attractive bar charts with small error bars, suggesting low variability in the data, or implying great technical skill and accuracy of measurement on the part of presenter; but it is incorrect: the standard deviation of the means from several samples is the S.E.M.

Example: Suppose you want to find out the mean value of the resting membrane potential in human intercostal muscle fibres. Ethical approval and informed consent are obtained and intracellular recordings are made from muscle biopsies from three male subjects, matched for age, ethnicity and general health. Each biopsy contains hundreds of muscle fibres: you record from (ie sample) 30 fibres in each. Thus, 90 muscle fibres are sampled in total, from 3 muscles. Imagine the mean values of resting membrane potential and their standard deviations in each muscle preparation were as follows:

Muscle 1 \((n=30)\) : \(-72 \pm 25\) mV (estimated S.E.M. : \(25/\sqrt{30} = 4.6\))
Muscle 2 \((n=30)\) : \(-78 \pm 28\) mV (estimated S.E.M. : \(28/\sqrt{30} = 5.1\))
Muscle 3 \((n=30)\) : \(-67 \pm 23\) mV (estimated S.E.M. : \(23/\sqrt{30} = 4.2\))

Mean-of-Means : -72.3 mV
S.E.M. \((N=3)\) : ± 5.5 mV

This S.E.M is the standard deviation calculated from the mean values of 3 muscles, each obtained from recordings of 30 fibres. The S.E.M. is NOT \(5.5/\sqrt{3} = 3.18\); and certainly not \(5.5/\sqrt{90} = 0.58\!\)!

FAQ: When should I use the following methods to graphically present data?
a) Histogram?
To display the distribution of data values in a set of quantitative, continuous data.
b) Box-whisker plot?
To summarise the distribution of continuous data, especially data that are not normally distributed.
c) Bar chart?
To summarise the frequency or mean values in quantitative or categorical data.
Biologists commonly perch standard deviation/error bars on top of bar chart data. Such "dynamite plunger" plots are frowned on by statisticians, who argue it is more informative to used box-whisker plots to convey the variability in a data set.
d) Scatter plot?
To display the relationship between two quantitative variables.
e) Line graph?
To summarise the relationship between two quantitative, continuous variables where the value of the dependent variable (y-axis, ordinate) varies as a function of the independent variable (x-axis, abscissa).
f) Cumulative histogram?
When you would like to be able to quickly compare the fraction (frequency) of a population above or below a given measured value.

See, for example, for illustration of these methods and the types of data for which they are appropriate:.
2. STATISTICAL TESTING
Comparisons of mean values and their standard errors form the basis of a variety of statistical tests, used to evaluate the probability that measured differences could be due to random errors in measurement (i.e., chance). Statistical tests that assume measurements come from a population(s) with a Gaussian distribution include the Student t-test, for comparing two groups, and Analysis of Variance (ANOVA) for comparing more than two groups (see below). Such tests are designed to evaluate the "null hypothesis", that there is no difference between two groups of data. However, if the data are not normally distributed, or are "non-parametric" (that is, they are not continuous variables) then you mustn’t use t-tests or ANOVA. Fortunately, there are tests that are suitable for non-parametric data including the Mann-Whitney Test (non-parametric equivalent of the t-test) and the Kruskal-Wallis Test (non-parametric equivalent of ANOVA).

FAQ: What is the difference between a parametric and a non-parametric test?
Parametric tests evaluate the significance of differences on the basis of specific assumptions about the way the data are distributed. They usually assume that this is a normal distribution or a continuous, quantitative variable.

Many kinds of biological data are not normally distributed. Non-parametric tests are useful for these since they make no assumptions about distribution: they are "distribution-free", and therefore safer to use if you are unsure about the distribution characteristics. However, if you know or strongly suspect that the measure is a normally-distributed continuous variable, then the appropriate parametric test is likely to be more sensitive and to give a result in which you can have a greater degree of confidence (i.e., parametric tests have more power to rule out Type II error).

FAQ: What is statistical significance?
Statistical significance is an arbitrary judgement, based on the probability that differences observed could have arisen by chance. What the statistical tests tell us is the likelihood that measured differences can be attributed to normal variation. The result is usually expressed in the form of a "P" value. This is a measure of the probability that the mean values are actually drawn from the same single population and the differences are due to chance. In biological studies, statistical significance is normally defined as P<0.05; that is, the probability that the observed differences could have come about by chance is less than 1 in 20 (in other words less than 5%). This is also the probability of a Type I error (α), incorrectly rejecting a null hypothesis. (By analogy, if you flip a coin 20 times and observe that it comes up ‘heads’ 19 of them, you might not completely rule out a run of luck but you may naturally suspect that something fishy is going on.) As a rule of thumb, sample values (or means) that fall outside 2σ are regarded as “significant”. If the value falls outside 3σ (P<0.01) the difference is considered “highly significant”.

FAQ: When should I use the following statistical tests?
a) unpaired Student t-test?
With small samples (<30) of normally distributed (parametric) data.

b) paired Student t-test?
With small samples of data representing repeated measures (e.g., before-after) of the same parametric variable.
c) Mann-Whitney test?
With non-parametric (e.g. categorised; semi-quantitative) data; or continuous data that are not normally (gaussian) distributed.

d) Wilcoxon test?
With repeated (paired) measures of non-parametric data.

e) Analysis of Variance (ANOVA)?
Where there are more than two groups of the same type of data. Analysis of variance compare the variability within a group against the variability between groups. It is not correct to use repeated t-tests for example on data in multiple groups. A "post-hoc" test is used to establish which data are significantly smaller or larger with a positive outcome to ANOVA.

f) Tukey’s, Bonferroni, or Dunnett’s tests?
These are three examples of "post-hoc" tests that can be applied after using ANOVA to establish that the between group variance is greater than the within-group variance. Post-hoc tests find out which means are significantly different among the different groups of data. Tukey’s compares all groups of data; Bonferroni’s test takes one of the groups as the reference (e.g. control) and tests the other group means against this reference group. Dunn’s test is similar to Bonferroni’s but is less conservative and can lead to more ‘false positives’ but fewer ‘false negatives’.

g) Kruskal-Wallis test?
Where there are more than two groups of the same type of non-parametric data. Dunn’s test may be used post hoc to establish which groups are significantly different.

h) Pearson’s correlation?
When measuring the degree of correlation between two quantitative, continuous variables which individually, are normally distributed.

i) Spearman’s correlation?
When measuring the degree of correlation between discrete or categorical data. The test is non-parametric.

j) Chi-squared (χ²) test?
When comparing distributions of categorical data, working with the actual numbers (rather than percentage summaries). The test is non-parametric.

k) Kolmogorov-Smirnov test?
Also when evaluating differences between two distributions (rather than differences between their means). The K-S test is non-parametric (distribution-free).

FAQ: What is statistical power?
The power of a statistical test is the probability that the test will reject a false null hypothesis (that it will not make a Type II error, β). As power increases, the chances of a Type II error decreases. Power (1-β) increases with the reliability of the measurements (small standard deviation) and with the sample size. The graph below shows hypothetical distributions of a variable we wish to measure assuming they are different. The graph in red might be the distribution of results in our control group, which we may know about already from previous work; and the one in blue is the conjectured distribution if there is difference between them. Suppose we make measurements from a “blue” sample and find the mean lies in the zone marked “α” (whose cutoff is approximately at zα=0.05 level of the red distribution). We will reject the null hypothesis and conclude that the blue mean is different from the red.
However, from the shape and overlap of the distributions we can see there is a chance we are wrong and that our blue sample is simply equivalent to sampling from the extreme end of the red population. That is, we would be making a Type I error.

Now let’s suppose that the measurements we make in our experimental group produce a mean within the zone marked $\beta$. This time we accept the null hypothesis (the result is to the right of the significance cut-off level of $\alpha$) and conclude that there is no difference between the blue and red distributions even if there is. We are making a Type II error. In order to reduce this error, what we need to do is find a way of sharpening the curves sufficiently to maximise $(1-\beta)$.

There are basically three things we can do: reduce the variability in our estimates of the mean (i.e. reduce the standard error); relax our criterion for $\alpha$ (eg consider $P<0.1$ as the cutoff); or measure something else instead (increase the difference between the means, $\mu_1-\mu_2$, or reduce $\sigma_1$ or $\sigma_2$). The last option may not be available; we don’t want to relax our significance criterion because that increases our chances of making a Type I error. This leaves reducing the variability in our estimates of the mean, that is: reduce the standard error of the mean. Since $\text{SEM}=\frac{\sigma}{\sqrt{N}}$, then we reduce the SEM by increasing the number of experiments. Or put another way, increasing the number of experiments increases the power of the study $(1-\beta)$, and our ability to detect differences. If we wish to achieve a power of $(1-\beta)=0.8$, the sample size equation tells us how many experiments to do.

**FAQ: When planning a study, what should I take into account?**

- How much scatter you expect
- How willing you are to risk mistakenly finding a difference by chance
- How big a difference you are looking for
- How sure you need to be that your study will detect a difference, if it exists: In other words, how much statistical “power” you need
FAQ: How do I calculate an appropriate sample size?
In other words, how many experiments should I do if I want to be sure that there is or there is not a difference in the mean value of the measurements I am making from an experimental group compared with a control group?

\[ N \leq \left( \frac{\Phi^{-1}(1 - \beta) + z_{\alpha}}{\mu / \sigma} \right)^2 \]

The equation above is one of those you are unlikely to use directly, since it is built into sample-size calculators in several commercial or free online software packages or macros. But it is important to understand the reasoning behind it, which aims to minimise the number of experiments required to achieve confidence in experimental measurements. Apart from being generally good science, calculating the numbers of experiments is required when important ethical issues, such as the use of human subjects or animals, are involved. The equation tells us how many experiments we need to do in order to minimise the chances of a false conclusion that there is no difference between two groups of data, when in fact there is a difference. The variables are: the cumulative distribution function \( \Phi \) calculated from the equation for a Gaussian distribution; the upper limit of significance \( z_{\alpha} \), normally equivalent to \( P=0.05 \), the mean \( \mu \) and standard deviation \( \sigma \) of the measurements from the experimental group; and finally the statistical “power” of the experiment \( 1-\beta \). Statistical power is the probability of rejecting the null hypothesis when it is false. A n approximation of the equation is \( N = 16(\sigma/\mu)^2 \) which works well for many studies.

If we carry out an experiment comparing measurements from two groups, calculate the means and standard errors, then run the appropriate statistical test and discover \( P<0.05 \), we will then reject the null hypothesis and say the difference between the means is significant. However, we must conclude this even if there really is no difference. In such cases we would be making a “Type I” error. (An extreme example would be to quote Oscar Wilde, in The Importance of Being Earnest: “To lose one parent, Mr Worthing, may be regarded as a misfortune; to lose both looks like carelessness.” That’s a judgement call with a distinct possibility of Type I error.) On the other hand, if \( P>0.05 \) we will accept the null hypothesis and conclude there is no difference. We must do this even if the means really are from different populations. In that case we would be making a “Type II” error. The sample size calculation aims to reduce our chances of making a Type II error. Typically, in biological or biomedical experiments we aim for a cutoff power of \( (1-\beta)=0.8 \), at a significance level \( z_{\alpha}=0.5 \) (i.e. probability of a Type I error set at \( P<0.05 \)). Our ability to detect differences then rests with the size of the effect we are looking for \( (\Delta \mu) \) and the variability in the population(s) we are examining (their standard deviation, \( \sigma \)).

FAQ: What is meant by a “double-blind study”?
A way of avoiding investigator bias, in which neither the tester nor the recipient knows what treatment or group is being tested, until a code is broken at the end of the study. Double-blind studies are especially important in any studies involving an element of subjective appraisal, where investigators may run the risk of (even unconsciously) biasing their handling of animals or interpretation of the data; or where the outcome may critically influence the interpretation of another part of a wider study.
3. ANALYSIS OF VARIANCE

ONE-WAY ANOVA
One-way ANOVA is used to test for differences among two or more independent groups. Typically, however, the one-way ANOVA is used to test for differences among at least three groups, since the two-group case can be covered by a t-test (Gossett, 1908). Two-way analysis of variance is an extension to the one-way analysis of variance. There are two independent variables (hence the name two-way). Hence two-way ANOVA, also called two-factor ANOVA, determines how a response is affected by two factors. For example, you might measure a response to three different drugs in both men and women. Drug treatment is one factor and gender is the other.

Assumptions
* The populations from which the samples were obtained must be normally or approximately normally distributed.
* The samples must be independent.
* The variances of the populations must be equal.

Hypotheses
For One-way ANOVA the null hypothesis will be that all population means are equal, the alternative hypothesis is that at least one mean is different.

There are three sets of hypothesis with the two-way ANOVA. The null hypotheses for each of the sets are as follows:

1. The population means of the first factor are equal. This is like the one-way ANOVA for the row factor.
2. The population means of the second factor are equal. This is like the one-way ANOVA for the column factor.
3. There is no interaction between the two factors. This is similar to performing a test for independence with contingency tables.

Factors
The two independent variables in a two-way ANOVA are called factors. The idea is that there are two variables, factors, which affect the dependent variable. Each factor will have two or more levels within it, and the degrees of freedom for each factor is one less than the number of levels.

Treatment Groups
Treatment Groups are formed by making all possible combinations of the two factors. For example, if the first factor has 3 levels and the second factor has 2 levels, then there will be 3x2=6 different treatment groups.

http://people.richland.edu/james/lecture/m170/ch13-2wy.html
TWO-WAY ANOVA
Two-way analysis of variance is an extension to the one-way analysis of variance. There are two independent variables (hence the name two-way). Hence two-way ANOVA, also called two-factor ANOVA, determines how a response is affected by two factors. For example, you might measure a response to three different drugs in both men and women. Drug treatment is one factor and gender is the other.

Assumptions
* The populations from which the samples were obtained must be normally or approximately normally distributed.
* The samples must be independent.
* The variances of the populations must be equal.
* The groups must have the same sample size.

Hypotheses
There are three sets of hypothesis with the two-way ANOVA.

The null hypotheses for each of the sets are as follows:
1. The population means of the first factor are equal. This is like the one-way ANOVA for the row factor.
2. The population means of the second factor are equal. This is like the one-way ANOVA for the column factor.
3. There is no interaction between the two factors. This is similar to performing a test for independence with contingency tables.

Factors
The two independent variables in a two-way ANOVA are called factors. The idea is that there are two variables, factors, which affect the dependent variable. Each factor will have two or more levels within it, and the degrees of freedom for each factor is one less than the number of levels.

Treatment Groups
Treatment Groups are formed by making all possible combinations of the two factors. For example, if the first factor has 3 levels and the second factor has 2 levels, then there will be 3x2=6 different treatment groups.

http://people.richland.edu/james/lecture/m170/ch13-2wy.html
4. CORRELATION AND REGRESSION

CORRELATION
Correlation quantifies the degree to which two variables are related. It makes no assumptions about cause-and-effect. Correlation does not fit a line through the data points. You simply are computing a correlation coefficient \( r \) that tells you how much one variable tends to change when the other one does. When \( r \) is 0.0, there is no relationship. When \( r \) is positive, there is a trend that one variable goes up as the other one goes up. When \( r \) is negative, there is a trend that one variable goes up as the other one goes down.

Correlation and regression are not the same. Regression finds the best line that predicts \( Y \) from \( X \).

\[ r = \frac{1}{n - 1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{\sigma_x} \right) \left( \frac{Y_i - \bar{Y}}{\sigma_y} \right) \]

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>−0.09 to 0.0</td>
<td>0.0 to 0.09</td>
</tr>
<tr>
<td>Small</td>
<td>−0.3 to −0.1</td>
<td>0.1 to 0.3</td>
</tr>
<tr>
<td>Medium</td>
<td>−0.5 to −0.3</td>
<td>0.3 to 0.5</td>
</tr>
<tr>
<td>Strong</td>
<td>−1.0 to −0.5</td>
<td>0.5 to 1.0</td>
</tr>
</tbody>
</table>
Spearman’s Correlation Correlation coefficient (non parametric):

\[ \rho = \sum_{i=1}^{n} \frac{(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}} \]

A Spearman correlation of 1 results when the two variables being compared are monotonically related, even if their relationship is not linear. In contrast, this does not give a perfect Pearson correlation. The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of both samples.
LINEAR REGRESSION

This is a modeling of the relationship between a scalar variable $y$ and one or more explanatory variables denoted $x$

*Linear Least Squares method:* This most common form of curve fitting by linear regression minimizes the sum of the squares of the residuals (difference between observed $x_i$ and expected $x'_i$ values) for a relationship $y=mx+c$; where $m$ is the slope and $c$ is the intercept on the $y$ axis. In this formulation, the slope $m$, for example is given by:

$$m = \left( \frac{1}{n} \sum x_i x'_i \right)^{-1} \left( \frac{1}{n} \sum x_i y_i \right)$$

Although Correlation and linear regression are not the same, they can be related to one another. The square of the sample correlation coefficient, $r^2$, which is also known as the **coefficient of determination**, estimates the fraction of the variance in $Y$ that is explained by $X$ in a simple linear regression.

$$r(Y, \hat{Y})^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

where $\hat{Y}$ is the fitted value in the regression
NON-LINEAR REGRESSION
Nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. There are several methods: the names (like Runge-Kutta, for instance) and details need not concern us but here are some examples of non-linear curves.

HYPERBOLIC GROWTH
Enzyme kinetics typically follow a hyperbolic Michaelis-Menten form:

\[ v = \frac{V_{\text{max}} [S]}{K_m + [S]} \]
**EXPONENTIAL GROWTH OR DECAY**

This occurs when the growth rate of the value of a mathematical function is proportional to the function's current value.

\[ N(t) = N_0 e^{kt} \quad --- \text{Exponential Growth} \]

or

\[ N(t) = N_0 e^{-kt} \quad --- \text{Exponential Decay} \]

where \( k \) is the “time constant”, for example

A useful quantity when measuring decaying exponential functions is the half-life or “half-decay” time. The classic example of this in physics is the half-life of a radioactive isotope, whose rate of decay depends on the amount present. The repolarisation time for synaptic potentials is also usually expressed in terms of their half-decay time.

Measuring the time from one rate of decay to half that value is related to the time constant by the equations:

\[ N\left(t_{1/2}\right) = \left(\frac{1}{2}\right)N_0 \]

and

\[ k = \frac{\ln(2)}{t_{1/2}} \]
SIGMOID GROWTH
Many natural processes, including those of complex system learning curves, exhibit a progression from small beginnings that accelerates and approaches a climax over time. When a detailed description is lacking, a sigmoid function is often used. A sigmoid curve is produced by a mathematical function having an "S" shape. Often, sigmoid function refers to the special case of the logistic function:

\[ P(t) = \frac{1}{1 + e^{-t}} \]
5. SUMMARY

When planning to carry out a statistical analysis of data, remember the quotation by Oliver Cromwell, albeit in an altogether different context, “I beseech you…think it possible you may be mistaken”. Ask yourself:

- Sampling: is my sample size sufficient? Is it randomly drawn from the population?
- Are my measurements objective (eg “blinded”)
- Plot the raw data
- Decide what is the most appropriate summary plot
- Are the data parametric or non-parametric?
- Are there more than two groups of data to be compared?
- What is the probability that the differences observed could have arisen by chance?
- Is there a possibility that the apparent difference could be a “false positive” (Type I error)? What different kinds of experiment could or should be done to corroborate the data?
- Is there a possibility that the apparent absence of a difference is a “false negative” (Type II error)? How many more experiments of the same kind should be done?

Resources for measurement and graphing

Microsoft Excel: several graph types, regressions and curve fits
Graphpad Prism: Fairly comprehensive graphing, curve fitting and statistical analysis
SigmaPlot: Extensive scientific graphing and curve fitting
ImageJ: Public domain image processing package with built in measurement tools that can be adapted or programmed through macros and plug-ins to perform a wide range of measurement functions. Variants called NeuronJ and FiJi have added functionality for neuronal morphology
Matlab: Officiandos extensive mathematical and graphical programming/macro language with an extensive range of public domain code, if you can find it. Expensive and requires programming skills.

Links

Free online statistics tests and tools:
http://www.quantitativeskills.com/sisa/index.htm
http://statpages.org/
http://www.dssresearch.com/toolkit/spcalc/power_a2.asp

Type 1 and Type 11 Errors explained and a nifty Applet to play with to illustrate the influences and effects of increasing statistical power
http://intuitor.com/statistics/T1T2Errors.html

Statistics notes/tutorials others:
http://people.richland.edu/james/lecture/m170/

Graphpad Prism help:
APPENDIX I : DATA PROBLEM

Resting Membrane Potentials and MEPP Frequency in Mutant Mice.

Imagine you have undertaken experiments to investigate some neurophysiological properties of a new mutant mouse that is beginning to attract worldwide interest. The observable phenotype in the mutant indicates lethargy and liminally-discernible flaccid paralysis in the hind limbs. You make the following observations, summarized in Table 1 (next page).

In your first experiment you find that the resting membrane potential recorded in 20 muscle fibres (postsynaptic cells) in the flexor digitorum brevis (FDB) muscle from a normal (“wild-type”) mouse soleus muscle has a mean and standard deviation of -70± 3.2 mV. Increasing the extracellular potassium ion concentration from 5 mM to 20 mM and resampling a different population (but the same number) of fibres in the same muscle yielded a mean of -55 mV ± 2.1 mV . There was an apparent increase in frequency of spontaneous miniature postsynaptic potentials (“minis”) from 1.3± 1.2 events per second to 5.2 ± 2.4 per sec in the two solutions.

You make similar recordings the following week in FDB muscles from a mutant mouse. Initial resting potential was -82 ± 5.4 mV and in 20 mM K+ was -77 ± 3.5 mV and mini frequency was 0.84 ± 1.1 per sec in the 5 mM K+ control solution, to 1.2 ± 1.5 per sec in the test solution. In the time remaining on the project, you repeat the experiments in muscles from 2 further normal (wild-type) mice and 5 mutant mice (recordings from 20 postsynaptic fibres in each case). Examine the data in the table on the next page and answer the following questions:

a. Are any of the differences within and between the mice statistically significant?

b. Formulate a testable hypothesis that would account for the behaviour of the mouse and the electrophysiological data. Show quantitatively, from a consideration of the Goldman Equation and of other research literature, how the above data are consistent with your hypothesis. How might the hypothesis be tested experimentally?

c. Suppose evidence from your experiment supported your hypothesis. How would you investigate it further at systems, cellular and molecular levels?

d. What ethical issues might be raised by experiments required for this further research?
<table>
<thead>
<tr>
<th>Resting potential (mV)</th>
<th>Normal K</th>
<th>High K</th>
<th>Normal K</th>
<th>High K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse 1 (male)</td>
<td>-70±3.2</td>
<td>-55±2.1</td>
<td>1.3±1.2</td>
<td>5.2±2.4</td>
</tr>
<tr>
<td>Mouse 2 (male)</td>
<td>-65±2.8</td>
<td>-56±1.9</td>
<td>2.1±1.8</td>
<td>4.9±2.2</td>
</tr>
<tr>
<td>Mouse 3 (female)</td>
<td>-73±4.9</td>
<td>-60±3.1</td>
<td>0.8±0.4</td>
<td>3.7±2.5</td>
</tr>
<tr>
<td>Mouse 1 Male</td>
<td>-82±5.4</td>
<td>-77±3.5</td>
<td>0.84±1.1</td>
<td>1.2±1.5</td>
</tr>
<tr>
<td>Mouse 2 Male</td>
<td>-76±3.2</td>
<td>-65±2.7</td>
<td>1.2±0.3</td>
<td>1.9±0.6</td>
</tr>
<tr>
<td>Mouse 3 Male</td>
<td>-72±0.3</td>
<td>-66±1.9</td>
<td>0.5±0.2</td>
<td>1.1±0.4</td>
</tr>
<tr>
<td>Mouse 4 Female</td>
<td>-68±1.9</td>
<td>-54±3.2</td>
<td>1.7±1.0</td>
<td>4.2±1.3</td>
</tr>
<tr>
<td>Mouse 5 Female</td>
<td>-71±0.5</td>
<td>-58±4.1</td>
<td>1.9±0.7</td>
<td>6.1±1.9</td>
</tr>
<tr>
<td>Mouse 6 Female</td>
<td>-69±1.1</td>
<td>-54±3.5</td>
<td>2.2±1.2</td>
<td>5.1±0.5</td>
</tr>
</tbody>
</table>
APPENDIX II

Pragmatic Rules of Thumb:

<table>
<thead>
<tr>
<th>Overview</th>
<th>Parametric</th>
<th>Non-Parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>Mean ± SD</td>
<td>Median (Scattergraph)</td>
</tr>
<tr>
<td>2 Groups</td>
<td>Student's t-test</td>
<td>Mann Whitney U-test</td>
</tr>
<tr>
<td>3+ Groups</td>
<td>ANOVA</td>
<td>Kruskal Wallis</td>
</tr>
<tr>
<td>Post Hoc Test</td>
<td>t-test + Bonferroni</td>
<td>MW + Bonferroni</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>Pearson</td>
<td>Spearman</td>
</tr>
</tbody>
</table>