Paper TU01

Basic Statistics and Commonly Encountered Tests for Statistical Programmers

Maria Efstathiou, Quintiles, Bracknell, UK

ABSTRACT

This paper provides an overview of basic statistical principles and tests. The definition and interpretation of terms such as p-values, confidence intervals, Type I and Type II errors and power of tests are introduced. Commonly encountered tests and procedures, such as t-tests, analysis of variance, tests for binary and censored data are also discussed. Examples using SAS® code are provided, linking the theory with programmers' everyday experience.

INTRODUCTION

The purpose of this paper is to provide an introductory background for the statistical procedures and tests commonly used by programmers in their work and to provide a refresher for those who have already come across these ideas in the past. The emphasis will be on procedures using normal distribution theory, although the analysis of binary and survival data will be discussed briefly too.

Statistics is primarily a tool that enables us to make decisions in the presence of uncertainty. For example, given a clinical trial comparing results from patients taking a test drug to results from patients taking a dummy placebo, we cannot usually tell with certainty if the test drug is the same as placebo or not. We can, however, say whether it is likely to be the same or different. We do so by means of estimated risks, based on the laws of probability. In the following sections we shall give some basic definitions and see how these are used in clinical trials in order to decide on the efficacy or otherwise of new treatments for patients.

DEFINITIONS

In this section we shall provide the definitions for the most common statistics used, namely the sample mean, variance, standard deviation, standard error and confidence intervals. We shall then proceed to describe the framework for hypothesis testing and define Type I and Type II error, power and p-values.

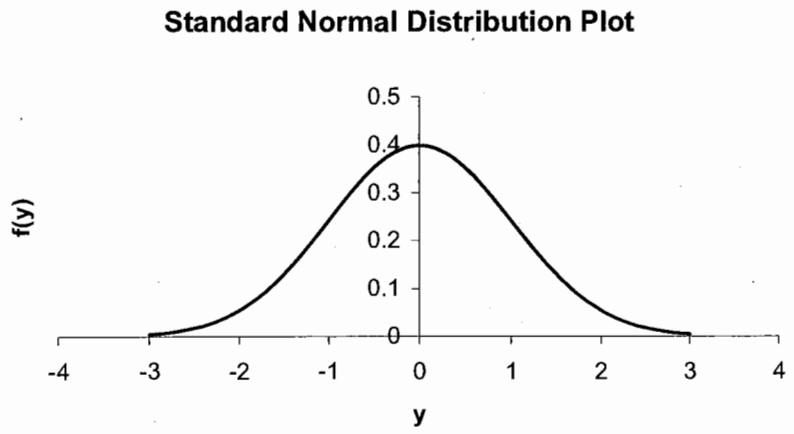
NORMAL DISTRIBUTION

We start by a quick reminder of the normal distribution and some associated terminology, as this will be used throughout the discussions in this paper. This is the most important probability distribution for a continuous variable. It is also known as a Gaussian distribution. The formula for this distribution is given by

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2} \right]$$ and y is said to follow the normal distribution with mean μ and variance σ^2.

Note that π is the constant 3.14159.... When the mean equals 0 and the variance 1, we have the special case of a *standard normal distribution* and a variable which follows the standard normal is usually denoted by z. Of special interest are the percentiles of the standard normal distribution. The k^{th} percentile of the standard normal distribution, usually denoted by z_k is the point for which $Pr(Z < z_k) = k$. In other words, it is the point of the distribution that leaves coverage area k to its left and $1-k$ to its right. E.g. the 0.95 (or 95%) percentile of the standard normal distribution is 1.64, whilst the 0.975 percentile is 1.96. There are tables in most basic statistical books which give the percentiles of the standard normal distribution. The shape of the standard normal distribution can be seen in Figure 1 below.

Figure 1



SOME BASIC STATISTICS

We shall now define very briefly some of the most common statistics that are used routinely in statistical analysis. Let us assume that we have a random sample of observations $y_1, y_2, \ldots, y_n$. The (sample) mean is then defined as $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, i.e. it is the average of the observations. This is a measure of the location or whereabouts of our distribution and it is the simplest and most common way of summarizing a number of observations assumed to come from the same distribution (population) using a single number. The sample variance of these observations is defined as $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$, whilst the standard deviation is simply defined as the square root of the variance, in other words $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$. Both the standard deviation and the variance are measures of the spread of our observations about their mean. Another statistic also very commonly presented in data summaries is the standard error. This is in fact the standard deviation of the sample mean ($\bar{y}$) and can be found by dividing the standard deviation by the square root of the number of observations, n, in the sample, i.e. $SE(\bar{y}) = s/\sqrt{n}$.

We shall see how these statistics are used in statistical inference in the sections that follow.

CONFIDENCE INTERVALS

A confidence interval is an interval which has a certain probability of containing the true value of a parameter of interest. For example, a 95% confidence interval for the difference in treatment mean blood pressure is an interval which contains the true treatment difference with probability 0.95. The width of the confidence interval indicates the precision with which we estimate the parameter in question. In the case of continuous data which can be assumed to follow the normal distribution, the confidence interval is defined as $\bar{y} \pm t * SE(\bar{y})$. The value of t depends on the sample size (number of observations) and the confidence level required. For a 95% confidence interval and provided that the number of observations is sufficiently large (30 or larger), t will be about 2.

HYPOTHESIS TESTING

Suppose we are testing a new blood pressure drug (test drug) which is hoped to be more effective than the current market standard (control drug). These are to be compared in a parallel group randomized clinical trial, the objective of which is to demonstrate the superiority of the test drug over the control. In order to assess whether the test drug is better than control, we usually start by assuming that the two drugs are in fact the same.

This assumption of no difference is known as the null hypothesis, commonly denoted by H_0. The alternative hypothesis is that the test drug is superior to control, and is denoted by H_1. Having conducted the trial, we then judge how likely it is to have observed the collected data under the assumption that the null hypothesis is true. If this

seems likely, we can conclude that the null hypothesis of no difference is reasonable. If it does not seem very likely, i.e. the data seem incompatible with the view that the treatments are the same, then we can conclude that the null hypothesis is not very reasonable, and accept the alternative hypothesis instead.

Given the above, we see that we can make one of two types of mistake. If the drugs truly are the same, we might mistakenly conclude that they are different. This is known as a Type I error. On the other hand, if the drugs are different, we might mistakenly conclude them to be the same. This is known as a Type II error. Clinical trials are typically designed to control the risk of these errors occurring. The risk of a Type I error is usually denoted as α, known as the "significance level" and this is often set at 5%. The risk of a Type II error, i.e. risk of concluding that the drugs are the same when they are not, is usually denoted by β. Common values for β are 10% or 20%. The power of our trial is then the probability of concluding that the treatments are different when they actually are. By definition, this is $1-\beta$ and it is the chance we have of detecting a difference between the two treatments, when a difference really exists. Hence we see that common values of power are 80% and 90%.

When designing a clinical trial, formulae are available that enable us to determine trial (sample) size in order to achieve a desired level of power, given assumptions about data variability and the magnitude of the treatment difference if indeed the treatments are not the same.

P-VALUE

The p-value is defined as the probability of observing data at least as extreme as the data we have observed assuming the null hypothesis is true. In the case of our blood pressure example, it is the probability of all possible trial outcomes where the treatment difference is greater than the difference we have observed, under the assumption that the treatments are the same. We will explore this more in an example to follow.

The p-value obtained from a statistical test is compared to the pre-defined significance level of the trial (e.g. 5%). If the p-value is smaller than the significance level, then we can reject the null hypothesis in favour of the alternative. If the p-value is higher than the significance level, then we conclude that the data observed is consistent with the null hypothesis, which is therefore accepted. It is important to note, however, that a p-value that is greater than the significance level does not prove that the null hypothesis is true, rather that we failed to see sufficiently strong evidence that it is not true. We must also bear in mind that the size of the p-value strengthens or weakens the evidence for or against the null hypothesis for the specific trial. For instance, a null hypothesis of 0.0001 provides strong evidence against the null hypothesis. On the other hand, borderline p-values like 0.049 or 0.051 can present difficulties in interpretation.

COMMONLY ENCOUNTERED STATISTICAL METHODS

In the sections that follow we shall present some examples of statistical tests that are commonly used for inference in clinical trials. We shall start by describing the use of the t-test as well as the more general analysis of variance methods, which can be applied whenever our data can be assumed to follow a normal distribution. We shall then consider two cases which do not follow such a distribution, namely binary data and survival data.

NORMAL DATA

1. *T*-TEST FOR A SINGLE MEAN

We can use this when we want to test whether the mean of a distribution has a particular value, under the assumption that the data we have collected are independent identically distributed and follow a normal distribution. The hypothesis can be set as follows:

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$.

We test this hypothesis by constructing the following statistic:

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \qquad (1)$$

This follows a *t* distribution with *n-1* degrees of freedom. We reject H_0 if $|t| \geq t_{1-\alpha/2}$, where $t_{1-\alpha/2}$ is the 1-α/2 percentile of the *t* distribution (defined similarly to the percentiles of the normal distribution mentioned earlier).

For example, let us assume that we are looking at a new blood pressure drug and would like to see what its effect is in changing (ideally lowering) blood pressure in subjects after a certain number of weeks of treatment. The response variable of interest is the change in systolic blood pressure from baseline after 4 weeks of treatment. Data for this example are given in Table 1 at the end of the paper.

Using the SAS procedure Proc MEANS, we can calculated the mean and standard deviation of the change in baseline corresponding to treatment 2, which in this case is the test drug. The code for this is given by

```
proc means data = test;
      where tmt = 2;
      var chg;
run;
```

and the resulting output from SAS is

The MEANS Procedure

Analysis Variable : chg

N	Mean	Std Dev	Minimum	Maximum
20	-2.8980000	2.3164665	-7.2900000	1.9500000

In the notation used above, we are trying to test the hypothesis H_0: $\mu = 0$, versus the alternative H_1: $\mu \neq 0$. For the calculation of the *t* statistic we have *n*=20, $\bar{y} = -2.898$ and s=2.3165. Applying these to formula (1) above, we obtain *t*=-5.59, with absolute value 5.59. The critical value of the *t* distribution with 19 degrees of freedom at the 5% level can be found from statistical tables as $t_{0.975} = 2.09$. As 5.59 is clearly larger than the critical value 2.09, we can conclude that there is strong evidence that the average change from baseline in systolic blood pressure is not 0. In fact, given that the observed mean is negative, we have evidence that the new drug on average reduces systolic blood pressure.

2. *T*-TEST FOR THE COMPARISON OF TWO INDEPENDENT MEANS

The test described above for a single mean of interest can easily be extended to the comparison of two means as follows. Suppose that we want to compare the change from baseline between the two treatment groups, the test drug and the placebo group. The data are again found in Table 1, where, now, we are also interested in the change from baseline observed in treatment group1, i.e. the placebo group. The null and alternative hypotheses can now be stated as

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$.

Notice that for the purposes of the example, we have chosen the two-sided alternative hypothesis, although the one-sided alternative ($\mu_1 > \mu_2$) would also have been an appropriate choice. The t-statistic in this case takes the form

$t = \dfrac{\bar{y}_1 - \bar{y}_2}{SE(\bar{y}_1 - \bar{y}_2)}$. $\bar{y}_1, \bar{y}_2$ are the sample means for the two treatment groups and $SE(\bar{y}_1 - \bar{y}_2)$ is the standard error of the difference of the means, given by $SE(\bar{y}_1 - \bar{y}_2) = \sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$, with n_1, n_2 the number of observations in each treatment group, $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$ and s_1^2, s_2^2 the sample variances of the two groups, calculated as described in the section "Some basic statistics" earlier. Using Proc MEANS again, this time for both treatment groups, we can obtain

tmt=1

The MEANS Procedure

Analysis Variable : chg

N	Mean	Std Dev	Minimum	Maximum
20	-0.4405000	2.3746678	-6.1700000	3.5000000

tmt=2

Analysis Variable : chg

N	Mean	Std Dev	Minimum	Maximum
20	-2.8980000	2.3164665	-7.2900000	1.9500000

In the notation used above, we have $n_1 = n_2 = 20, \bar{y}_1 = -0.441, \bar{y}_2 = -2.898, s_1^2 = 5.6390, s_2^2 = 2.3165$, resulting in a value of the t-statistic of t=3.31. The critical value for rejection of the null hypothesis at the 5% level is given by the 97.5 percentile of the t-distribution with (20-1)+(20-1) = 38 degrees of freedom, and can be found from statistical tables to be $t_{0.975}$ = 2.02. The value 3.31 that we obtained as our test statistic is comfortably higher than the critical value 2.02, which provides evidence for us to reject the null hypothesis of no difference between the two means and accept the alternative that there is a difference between them.

The discussion above has involved some manual calculations of the test statistics and critical value. We can in fact carry out this analysis in SAS, and there are a number of different procedures that allow us to do this. Here we shall illustrate the example by using just two of these, namely Proc TTEST and Proc GLM.

ANALYSIS USING PROC TTEST

The code and corresponding output from proc TTEST are given below:

```
proc ttest data=test;
  class tmt;
  var chg;
run;
```

The TTEST Procedure

Statistics

Variable	tmt	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
chg	1	20	-1.552	-0.441	0.6709	1.8059	2.3747	3.4684	0.531
chg	2	20	-3.982	-2.898	-1.814	1.7617	2.3165	3.3834	0.518
chg	Diff (1-2)		0.9558	2.4575	3.9592	1.9171	2.3457	3.0231	0.7418

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
chg	**Pooled**	**Equal**	**38**	**3.31**	**0.0020**
chg	Satterthwaite	Unequal	38	3.31	0.0020

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
chg	Folded F	19	19	1.05	0.9150

The row of interest for the hypothesis testing just discussed is the one highlighted in bold, using the "pooled" method, which assumes that the two treatment groups have a common, but unknown variance. The *t*-value 3.31 is the same as the one we obtained with our manual calculations. The Pr > |t| column is the p-value, which, in this case, is found to be 0.0020. This is well below the 0.05 (5%) significance level, so we can comfortably conclude that the null hypothesis of no difference between the two treatment groups does not seem likely and that we can accept the alternative hypothesis instead. Note that the mean difference between the two groups is positive (2.46), which implies that the reduction in blood pressure is greater from treatment 2 (i.e. the test drug).

ANALYSIS USING PROC GLM

The same analysis can be carried out using proc GLM and we give the code and corresponding output below.

```
proc glm data=test;
  class tmt;
  model chg=tmt;
run;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
tmt	2	1 2

Number of observations 40

Dependent Variable: chg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	60.3930625	60.3930625	**10.98**	**0.0020**
Error	**38**	209.0962150	5.5025320		
Corrected Total	39	269.4892775			

R-Square	Coeff Var	Root MSE	chg Mean
0.224102	-140.5270	2.345748	-1.669250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
tmt	1	60.39306250	60.39306250	10.98	0.0020

Source	DF	Type III SS	Mean Square	F Value	Pr > F
tmt	1	60.39306250	60.39306250	**10.98**	**0.0020**

Even though it seems that lots of different numbers have appeared in the output of the GLM procedure, it is actually equivalent to the results of the t-test described above. The F-value that is highlighted in bold is in fact the square of the t-value that we obtained from Proc TTEST above. (The F distribution is a more general family of distributions of which the t-distribution discussed previously is a special case). The Pr > F column gives the p-value, which again is 0.0020 as from the t-test before. Note the "tmt" and "model" rows of the output give exactly the same results. This is because there is no other term in the model that we use apart from the treatment groups. If we wanted to investigate the influence of other covariates or factors in the performance of the two treatment groups (e.g. systolic blood pressure at baseline or smoking habits of patients), then we could do this using proc GLM and in that case, we would be getting different pieces of information in each of the aforementioned rows. We shall discuss this very briefly in the paragraph below.

ANALYSIS OF VARIANCE

The *t* test described above is the simplest case of an analysis of variance, applicable when we are interested in comparing the means of two treatment groups only and no other factors are considered to influence the difference between the treatment means. This can be extended to the simultaneous comparison of more than two treatment groups. In this case, the one-way analysis of variance can be applied and give us an answer on whether overall there is evidence of a difference between the various treatment means. We can do this in SAS using Proc GLM in exactly the same way as we did above, again checking the "tmt" row in order to find the p-value for the overall treatment effect.

A further addition to the investigations might be to examine the influence of other factors, not directly related to treatment, but which are thought to influence the outcome of interest. For instance, in the example we have discussed above, it would be very common to investigate the effect of the baseline value of systolic blood pressure in the change from baseline observed and analysed at the end of the trial. This would be very useful as it is possible that, by chance, the average baseline value was greater in the placebo patients compared to the test drug patients. With analysis of variance we are able to "remove" the influence of these factors prior to carrying out the comparison of the treatments. For instance, if we think that whether the patient is a smoker or not can affect the outcome of a drug for treating stomach ulcer, then we may want to include this information in the model we are going to use for analyzing the data.

Finally, analysis of variance allows us to examine the effect of different doses of the same treatment, when we consider dose as a continuous variable itself, and the effect obtained is assumed to be linearly related to the dose used. This is known as "linear regression" or "dose response analysis" and can be seen as a special case of analysis of variance.

NON-NORMAL DATA

The discussion so far has focused on the case of continuous data which can be assumed to follow a normal distribution. Other types of data commonly encountered in the context of clinical trials include binary data, counts and survival data. Whilst in some cases it may be possible to assume that these data can still be reasonably approximated by a normal distribution, methods that are particularly suited for these types of data have been developed and we shall briefly discuss some of these below.

ANALYSIS OF PROPORTIONS

Let us now assume that instead of analyzing the change from baseline in systolic blood pressure, the outcome of interest for a clinician was to establish whether the test drug is "successful" or not and that the "success" of the drug was measured by whether a patient achieved a reduction of at least 2 units in systolic blood pressure compared to the baseline.

This is a case of a response that is binary in nature (yes/no, responder/non-responder, success/failure) and one way to approach the problem of deciding on whether the test drug is better than the placebo is to compare the proportion of responders (successes), in the test drug group, to that of the responders in the control group.

There are various ways in which we can do this and we shall start by considering the simplest one. Let r_1 be the number of successes in the test drug group, which has n_1 patients and r_2, n_2 the number of successes and patients in the control group. Define $p1 = r_1/n_1$ and $p_2=r_2/n_2$ to be the observed proportions of successes in the two groups. We wish to test the hypothesis that there is no difference in the proportion of responders between the two groups, versus the (two-sided) alternative hypothesis that there is a difference. Let $p=(r_1+r_2)/(n_1+n_2)$ denote the proportion of successes across the two groups and $q=1-p$. We can then define the statistic

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

which follows approximately a standardized normal distribution. Similarly to what we did with the t-test, we can reject the null hypothesis in favour of the alternative if $|z| > z_{1-\alpha/2}$, where $z_{1-a/2}$ is the $1- a/2$ percentile of the standard normal distribution.

Given the data in Table 1, where now we are focusing on the "responder" column, and a success is denoted by "Y" we have

$p_1 = 0.2$, $p_2 = 0.7$, $p_1-p_2 = -0.5$, $p=18/40 = 0.45$, $q=1-0.45=0.55$. With these, we obtain the value for the z statistic as

$z = \frac{-0.5}{\sqrt{0.45*0.55*0.1}} = -3.178.$ The absolute value of this (3.178) is to be compared with the 0.975 (=1-0.05/2) point of the normal distribution, which is 1.96. As3.178 is much larger than 1.96, and corresponds to a p-value of 0.0015, we can conclude that the null hypothesis of no difference in the proportion of responders between the two groups is not very likely and that we can accept the alternative hypothesis instead.

We can do this analysis in SAS easily using Proc FREQ. The code and the corresponding output are given below.

```
proc freq data=test;
  tables tmt*resp / chisq;
run;
```

The FREQ Procedure

Table of tmt by resp

tmt	resp		
Frequency Percent Row Pct Col Pct	N	Y	Total
1	16	4	20
	40.00	10.00	50.00
	80.00	**20.00**	
	72.73	22.22	
2	6	14	20
	15.00	35.00	50.00
	30.00	**70.00**	
	27.27	77.78	
Total	22	18	40
	55.00	45.00	100.00

Statistics for Table of tmt by resp

Statistic	DF	Value	Prob
Chi-Square	**1**	**10.1010**	**0.0015**
Likelihood Ratio Chi-Square	1	10.6004	0.0011
Continuity Adj. Chi-Square	1	8.1818	0.0042
Mantel-Haenszel Chi-Square	1	9.8485	0.0017
Phi Coefficient		0.5025	
Contingency Coefficient		0.4490	
Cramer's V		0.5025	

Fisher's Exact Test

Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.9998
Right-sided Pr >= F	0.0018
Table Probability (P)	0.0017
Two-sided Pr <= P	0.0036

Sample Size = 40

We can see that the procedure produces the results for many more tests than the one we have just discussed, highlighted in bold in the output above, and again in a slightly different form, as this is presented here as the square of the normal result that we obtained manually above. The percentages that we have highlighted in the table of treatment by responder are the ones we used in our calculations. While this example illustrates how we can approach such problems, the method chosen here is not necessarily the best for the particular data that we have. This is because in using this method, we have assumed that the ***z*** statistic that we obtained approximately follows a normal distribution. In cases of large sample sizes (n_1 and n_2) and where all four "cells" of the table have counts that are quite large (expected count larger than 5 is sometimes used as a rule of thumb), this assumption is reasonable. In the case that we have here, it is does not appear to be so, and it might be best to consider using another method,

such as Fisher's exact test, which, for small sample sizes, or for cases where we may have small proportions, gives much more accurate results. Indeed, we can see that the two-side p-value from Fisher's exact test is 0.0036, nearly two and a half times the one we obtained using the crude normal approximation. The continuity adjusted Chi-square value is also a test that is more appropriate for this data set than the most basic approach available. Having said all of that, the signal in this case is so strong, that all methods agree in the conclusion: there is clear evidence that there is a difference in the proportion of responders between the two groups.

SURVIVAL ANALYSIS

Survival analysis is applied whenever the data under consideration are "time to event" type. This can be "time to death", in which case it literally is "survival analysis". They can also be "time to relapse" of a disease, "time to disease free status", etc. The fundamental difference between these types of data and the types discussed so far is what is known as "censoring". Because of the nature of the data, the event of interest will not always be observed in the timeframe of observation. For instance, if we are conducting a clinical trial and the event that we are interested in is the time to symptom relief, then we may not observe this event for all individuals in the trial. Those individuals for whom the event is not observed (e.g. the trial stops or they drop out before they become symptom free) are said to be censored.

Survival analysis data are usually far from symmetric and therefore the bell-shaped normal distribution cannot describe them in a satisfactory way. The biggest problem with using the methods discussed earlier for the analysis of survival data, is, however, censoring itself. Why is it a problem? If we ignore censoring and take the average of the observed times to event, then we will underestimate the true average.

As a result, other methods have been developed for examining and analyzing survival data. The Kaplan-Meier estimator is perhaps the one most commonly used and can be carried out in SAS using Proc LIFETEST. Another common approach to modeling survival data is the Cox proportional hazards model, which can be fitted in SAS using Proc PHREG. Whilst discussion of these methods is beyond the scope of the current paper, we felt that is was worth pointing out the basic characteristic of survival data (i.e. censoring) which have made the development of a different approach to their analysis essential. We should point out, however, that, while the actual methods of analysis are indeed different, the fundamentals are the same: with survival analysis methods we are still trying to test hypotheses and the decide of their plausibility based on the evidence provided by the data.

BIAS AND HOW TO AVOID IT

No introduction to basic statistical terms of relevance to clinical trials would be complete without a brief discussion of Bias. An essential feature of statistical contributions to the design of clinical trials and indeed any experimental setting is an understanding of how bias may be introduced, and steps that should be taken to avoid it.

There are various places in which bias can be introduced in a trial. Two of the most common sources are selection bias and assessor bias. Selection bias can occur when an investigator assesses a subject prior to allocating him/her to a specific treatment. For example, bias would arise if sicker patients were systematically assigned to the test drug rather than control. In order to avoid this, we run randomized trials. The allocation of subjects to treatments is then decided by the randomization code. (Note that for open-label randomized trials, it is best to use central randomization so that the investigator cannot deduce what the next patient will be getting, otherwise selection bias may still be a problem). Assessor bias can occur when the subject or the investigator are aware of the treatment the subject received. In these cases, their judgment of their experiences during the trial may be affected by the knowledge of the treatment and even their own response to treatment may be altered. This can be addressed via blinding, which in most cases is double – neither the investigator nor the subject know the treatment which the subject receives. A third type of bias is that of analysis bias. If the analysis choices are determined after the data have been collected and the trial unblinded, then we run the risk of choosing the analysis that is most favourable to the experimental drug. This is avoided by specifying the details of the analysis prior to unblinding the data. The expertise that statisticians bring to the discussions about the design and conduct of a trial and the insight that they can provide in possible sources of bias and ways to avoid it is indeed one of the areas where statistics adds value and contributes to a clinical trial of high quality.

CONCLUSION

The aim of statistics is to distinguish between the true signal (treatment effect) and the noise (random variation) present in the data and to help design studies that are free from bias. Normal distribution theory provides a powerful tool for statistical inference. Other distributions provide the theoretical background for non-normal data (binary, censored). Assumptions about the distribution of the data must be considered prior to designing a study and checked once the data have been collected. Understanding the principles behind the methods can greatly improve the quality of the work of a programmer and indeed their enjoyment of it.

TABLE 1
Data TEST
TMT = 1 corresponds to the placebo group
TMT = 2 corresponds to the test drug group

Subject (subjid)	Treatment (tmt)	Change from Baseline in Systolic BP (chg)	Responder (resp)
1	1	1.27	N
2	1	0.46	N
3	1	2.09	N
4	1	-0.82	N
5	1	-1.25	N
6	1	-6.17	Y
7	1	-3.6	Y
8	1	-0.97	N
9	1	-1.22	N
10	1	2.27	N
11	1	-1.81	N
12	1	0.4	N
13	1	-2.41	Y
14	1	3.5	N
15	1	0.45	N
16	1	1.95	N
17	1	-0.06	N
18	1	1.26	N
19	1	-0.04	N
20	1	-4.11	Y
21	2	-4.05	Y
22	2	-5.13	Y
23	2	-0.4	N
24	2	-5.49	Y
25	2	-2.23	Y
26	2	-2.61	Y
27	2	-6.58	Y
28	2	-1.87	N
29	2	0.69	N
30	2	-7.29	Y
31	2	-1.58	N
32	2	-2.68	Y
33	2	-3.41	Y
34	2	-0.53	N
35	2	-2.46	Y
36	2	-4.26	Y
37	2	-4.37	Y
38	2	-2.82	Y
39	2	-2.84	Y
40	2	1.95	N

The data above were simulated using the following SAS code

```
Data test (keep=tmt subjid chg resp);
  do tmt=1 to 2;
    do subj=1 to 20;
         chg=round(2*rannor(0), 0.01);
         subjid=subj;
         if tmt=2 then do;
              chg=chg-3;
             subjid=subj+20;
         end;
         if chg <= -2 then resp='Y';
         else resp='N';
         output;
       end;
  end;
run;
```

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Maria Efstathiou
Quintiles
Station House
Market Street
Bracknell RG12 1HX
Work Phone: +44 1344 708614
Fax: +44 1344 708106
Email: maria.efstathiou@quintiles.com