

What makes a “Statistical Programmer” different from a “Programmer”

PharmaSUG 2016 - Paper IB01

What makes a “Statistical Programmer” different from a “Programmer”

Arun Raj Vidhyadharan, inVentiv Health, Somerset, NJ

Sunil Mohan Jairath, inVentiv Health, Somerset, NJ

ABSTRACT

In clinical SAS® programming world programmers come from different backgrounds like engineering, biotech, biomedical etc. and everyone knows how to write a program as per specifications. At the same time some programmers who understand statistics can look beyond records and their statistical awareness helps them to understand data in a better way. Statistics is not about all crazy formulas, but it's a set of tools that helps us in understanding, analyzing and presenting clinical trial data. We all produce lot of tables and at the same time having basic understanding of statistics will be beneficial for us as we can contribute more than just creating TLFs. So next time when statistician asks to provide standard error instead of standard deviation; or statistician suggested using different formulae each time they calculate p-value we can also provide our input. The goal of this paper is to give the programmer insight to understand basic statistics involved in creation of tables, as a result, a programmer understand and check their numbers , and provide input to improve the quality of the final product and help company to save time and resources.

INTRODUCTION

In programming for clinical trials, both programmers and statisticians have distinct roles to play in the development of statistical reports used in regulatory filings. The statisticians design and implement statistical analysis using protocols that provide more specific information in a Statistical Analysis Plan (SAP). Based on a SAP the programmers create analysis files and reporting programs to produce the tables, figures, and listings required to complete a clinical study report. While the jobs of a programmer and statistician are separate, the most effective statisticians understand the implications of their requests on programs. So with a basic fundamental understanding of statistics, the programmer can provide valuable feedback to the statistician. Alternately, a lack of common understanding can lead to unnecessary work. The goal of this paper is to enable programmers to understand the basic statistics involved in producing an output and ask intelligent questions thereby reducing the amount of work required to complete the statistical analysis.

This paper will cover some basic real life questions that every statistical programmer would have faced and even searched internet at least once. Internet can bring you some statistical concepts with some complicated mathematical expressions, but the question is how we relate them with our statistical work.

CASE STUDY 1: P-VALUE

Understanding P-value:

Every statistical programmer comes across P-value tables as it is needed everywhere as far as any statistical testing is concerned.

To have a better understanding of P-value we will divide it into parts: Null hypothesis and Alternate Hypothesis

In any clinical trial the goal is to get information about target population using available sample population and sample size. The purpose can vary based on sponsor's interest for example:

Is study drug effective?

Is study drug better than any other drug involved in clinical trial?

The method to get answers to these questions is by “Testing a Hypothesis”

Testing of hypothesis should necessarily have two hypotheses statements at the very least. First is the null hypothesis, which is the “no change” statement, against the alternative hypothesis, “the change” statement. And most importantly, the alternative hypothesis should always support the purpose of the trial.

Consider, for example, an experimental new drug for lowering blood pressure. The researchers would not be able to administer the drug to every single person with hypertension (high blood pressure) around the world. So a sample of hypertensive patients would be selected for the study, perhaps from the researchers' local community. After administering the drug to an experimental group and a placebo treatment to a control group, the researchers would

calculate the mean changes in blood pressure for the groups over the study period. Let's say that the mean value for blood pressure decreased by 6.2 mm Hg and 1.2 mm Hg in the experiment and control groups, respectively. The difference in change between the two groups is 5.0 mm Hg. In repeated experiments on different samples of the population of hypertensive individuals in the world, we would not expect to observe a difference in change of exactly 5.0 mm Hg for systolic blood pressure. Instead, we would expect sampling variability.

To maintain objectivity, the null hypothesis and alternate hypothesis will be established prior to conducting their studies. Alternate hypothesis, is an informed statement (not a wild guess) about their true expectations and beliefs. Statistical tests help us determine whether the null hypothesis or the alternate hypothesis is more tenable. So for this example:

Null Hypothesis - Average improvement in BP is same for study drug and placebo

Alternate Hypothesis - Average improvement in BP is greater for study drug as compared to placebo

The decision is based on the significance of study results. In statistics, the term "significance" refers to the mathematical probability of obtaining the observed results if the same experiment were conducted many times on different samples of subjects in the population of interest. In conducting tests of statistical significance (such as t-tests and ANOVA), researchers answer this central question: If the null hypothesis were true in the population (that is, if there really is no difference between groups and no treatment effect), what is the probability of obtaining the results that we observed in our experiment?

P-Value:

Now the million dollar question is: Is the 5.0 mm Hg difference truly due to the effects of the drug, or is the difference simply due to chance factors? This is where P-Value comes in picture.

The statistical test would yield a P value, which is the probability of observing a 5.0 mm Hg or greater difference in systolic blood pressure in repeated experiments if the null hypothesis were true. Let's say that the statistical test revealed a P value of .99. Here's the interpretation: If it's really true that the drug has no effect on systolic blood pressure, we would observe a 5.0 mm Hg difference between experimental and control subjects in 99 out of 100 repeated experiments. If we obtained such a result so frequently, we would be confident that the null hypothesis is tenable and that the drug doesn't really reduce blood pressure.

What if the test revealed a P value of .01? Here's the interpretation: If it's really true that the drug has no effect on systolic blood pressure, we would observe a 5.0 mm Hg difference between experimental and control subjects in 1 out of 100 repeated experiments. If we obtained such a result so rarely under the assumption that the null hypothesis is true, we would have to doubt that the null hypothesis is tenable. We would thus accept the research hypothesis that the drug does reduce blood pressure. We would say that the 5.0 mm Hg difference is statistically significant. As you may have learned in statistics class, a more accurate interpretation is that the 5.0 mm Hg difference is statistically reliable

As the P value gets lower (i.e., closer to 0% and farther away from 100%), researchers are more inclined to accept the research hypothesis and to reject the null hypothesis. As you know, before beginning studies scientists conventionally establish cutoff points, which are called alpha values, to indicate what P values they will accept as significant. In many physiological studies, alpha values are set at .05 or .01. As you know, if the P values that are calculated in statistical tests are less than alpha—for example, if $P < .05$ —the researchers would conclude that their study results are statistically significant.

A relatively simple way to interpret P values is to think of them as representing how likely a result would occur by chance. For a calculated P value of .001, we can say that the observed outcome would be expected to occur by chance only 1 in 1,000 times in repeated tests on different samples of the population.

What's so special about 0.05?

It's about defining a boundary. The truth about the population remains unknown forever. So to decide based on probability would require a strict boundary and beyond that the null hypothesis would be rejected. The boundary can be considered as 0.01 if we want to make the test stricter. This boundary in probability is the level of significance for the test.

Can we decide that the null hypothesis is true if we get a high P-value?

No. We can only reject or not reject the null hypothesis. If the p-value comes greater than significance level, we can only say that we cannot reject the null hypothesis based on the given sample. Small p-values indicate very strong evidence against the null hypothesis, because it indicates that random variation was not responsible for the observed value of the sample. Large p-values indicate no evidence against the null hypothesis.

CASE STUDY 2: T-TESTS

What is T-TEST?

The **t-test** looks at the t-statistic, **t-distribution** and degrees of freedom to determine a p value (probability) that can be used to determine whether the population means differ. The **t-test** is one of a number of hypothesis tests. The greater the magnitude of T (it can be either positive or negative), the greater the evidence against the null hypothesis that there is no significant difference. The closer T is to 0, the more likely there isn't a significant difference.

Researchers developed cholesterol drug and gave it to 2 groups: After 1 month they noticed that group that received drug had mean cholesterol of 36 and group which received placebo has mean cholesterol of 34.

T-test tells if the difference is reliable or is it just a chance.

T test checks measures difference between the groups and compares it to difference within the groups. The T value is ratio of these 2 numbers: Variance between groups over variance within groups.

In our example for cholesterol drug the difference between mean cholesterol for 2 groups is 2 while difference within group is 6. So $2/6$ gives T-value of one third, which is not big enough. So how do we know if our difference is reliable? So with our t value has a corresponding P-value. And P-value tells us if the results are real or it's just a fluke.

When to Use

We can use t-tests in the following three situations;

- We want to test whether the mean is significantly different than a hypothesized value.
- We want to test whether means for two independent groups are significantly different.
- We want to test whether means for dependent or paired groups are significantly different.

However, to use a t-test at all, we must have interval variables that are assumed normally distributed.

How to implement in SAS?

To test whether the mean of one variable is significantly different than a hypothesized value, we can use the following SAS syntax:

```
PROC TTEST DATA= datasetname H0=hypothesizedvalue;  
    VAR variable_of_interest;  
RUN;
```

If we omitted the `H0=hypothesized value` option, SAS would use the default of `H0=0` when running the t-test.

In order to test whether the mean of two dependent groups are significantly different, we need to construct the SAS dataset in such a way that we have two observations per subject. We use the following slightly different SAS syntax:

```
PROC TTEST DATA= datasetname;  
    PAIRED dependent_variableA*dependent_variableB;  
RUN;
```

Testing whether the means of two independent groups are different is the most complicated type of t-test. For this type of t-test, we need to create a classification variable or dummy variable. Class variables are 0/1 binary indicator variables. An example of a class variable might be gender, where `gender=1` when the observation is male or `gender=0` when the observation is female. Another example could be a vital status variable that equals 1 when a person is alive and is 0 when the person is dead. Once a class variable has been created, we use the following SAS syntax to perform the desired t-test.

```
PROC TTEST DATA= datasetname;  
    CLASS classvariable;  
    VAR variable_of_interest;  
RUN;
```

What Is Degree of Freedom in T-TEST?

A simple (though not completely accurate) way of thinking about degrees of freedom is to imagine you are picking people to play in a team. You have eleven positions to fill and eleven people to put into those positions. How many decisions do you have? In fact you have ten, because when you come to the eleventh person, there is only one person and one position, so you have no choice. You thus have ten 'degrees of freedom' as it is called.

Likewise, when you have a sample, the degrees of freedom to allocate people in the sample to tests is one less than the sample size. So if there are N people in a sample, the degrees of freedom is N-1.

What does a negative t value mean in a T-TEST?

If it's a one-sample t-test, then a negative t-value only means that the sample mean was less than the test value.

If it's a two-sample t-test, then the negative t-value only means that the sample mean of the first population was less than the sample mean in the second population...the negative sign isn't all that important.

What's important is the p-value. Since your p-value was so small, you can reject the null hypothesis and conclude that the alternative is true.

Example for T-TEST:

We want to determine whether a weekend study session improves students' test scores. Six students are given a math test before the session, then they are re-tested after the weekend training. This is a matched pairs t-test, because the same subjects are being measured before and after some intervention.

Ho: $\mu_{\text{before}} = \mu_{\text{after}}$ Ha: $\mu_{\text{before}} \neq \mu_{\text{after}}$ Again, before we can analyze the data, we have to assume that data came from a normal distribution.

```
PROC TTEST DATA = study;
```

```
    TITLE "Example of Program for a Paired T-test";
```

```
    PAIRED before * after;
```

```
RUN;
```

The code tells SAS to do a paired t-test on the data set study, and it will compare the difference of the means between before and after.

```
Example of Program for a Paired T-test
The TTEST Procedure
Statistics
Difference      N      Lower CL      Mean      Upper CL      Lower CL      Std Dev      Std Dev      Upper CL      Std Err
before - after  6      -11.67    -7.333      -2.998      2.5787      4.1312      10.132      1.6865
T-Tests
Difference      DF      t Value      Pr > |t|
before - after      5      -4.35      0.0074
```

Interpreting output above:

The difference of the mean score (before-after) is -7.33; on average the scores before the weekend were lower than the scores after the training session.

Note: If in your paired statement you had typed "after*before" the average difference would be 7.33.

Is this difference statistically significant? To answer that question, look at the p-value. The t value for the test is -4.35, and the p-value is 0.0074.

If $\alpha = 0.05$, then the p-value < α , and we reject the null hypothesis. Therefore, we can conclude that average scores are different before and after the weekend session, and the training does improve test scores.

Case Study 3: BASIC SUMMARY STATISTICS IN A TABLE:

Parameter	Descriptive Statistics	Placebo (n=115)
Age (years)	N	115
	Mean (SD)	27.8 (3.59)
	Median	28.2
	Q1 - Q3	25 - 30
	Min - Max	19 - 41

Table 1: Descriptive Statistics for Age

What is Q1-Q3?

Quartiles are the values that divide a list of numbers into quarters. There are three quartiles: the first quartile (Q1), the second quartile (Q2), and the third quartile (Q3).

The first quartile (lower quartile, QL), is equal to the 25th percentile of the data. (splits off the lowest 25% of data from the highest 75%). The second (middle) quartile or median of a data set is equal to the 50th percentile of the data (cuts data in half). The third quartile, called upper quartile (QU), is equal to the 75th percentile of the data. (splits off the lowest 75% of data from highest 25%).

The table signifies that the average age of the 115 subjects belonging to Placebo group is approximately 28 (27.8) years with a deviation of 3.5 years.

1. The same is evident from the median age of 28 years, with half of the subjects below the median and remaining above it.
2. The Q1 and Q3 stand for the lower and upper quartiles of the data with 25% of the subjects lying below and above these values respectively, and the remaining 50% in between. So it is understood that 50% of the subjects belong to the 25-30 years age group.
3. The youngest of the subjects is 19 and the eldest is 41 with the rest in between.

Difference between SD and SE:

Standard Deviation and Standard Error are perhaps the two least understood statistics commonly shown in data tables. The following paper is intended to explain their meaning and provide additional insight on how they are used in data analysis. Both statistics are typically shown with the mean of a variable, and in a sense, they both speak about the mean. They are often referred to as the "standard deviation of the mean" and the "standard error of the mean." However, they are not interchangeable and represent very different concepts.

Standard Deviation

Standard Deviation (often abbreviated as "Std Dev" or "SD") provides an indication of how far the individual responses to a question vary or "deviate" from the mean. The SD is a measure of the dispersion of the data around the mean. SD tells the researcher how spreads out the responses are –

Are they concentrated around the mean, or scattered far & wide?

Did all of your respondents rate your product in the middle of your scale, or did some love it and some hate it?

Another way of looking at Standard Deviation is by plotting the distribution as a histogram of responses as shown in figure 2 below. A distribution with a low SD would display as a tall narrow shape, while a large SD would be indicated by a wider shape.

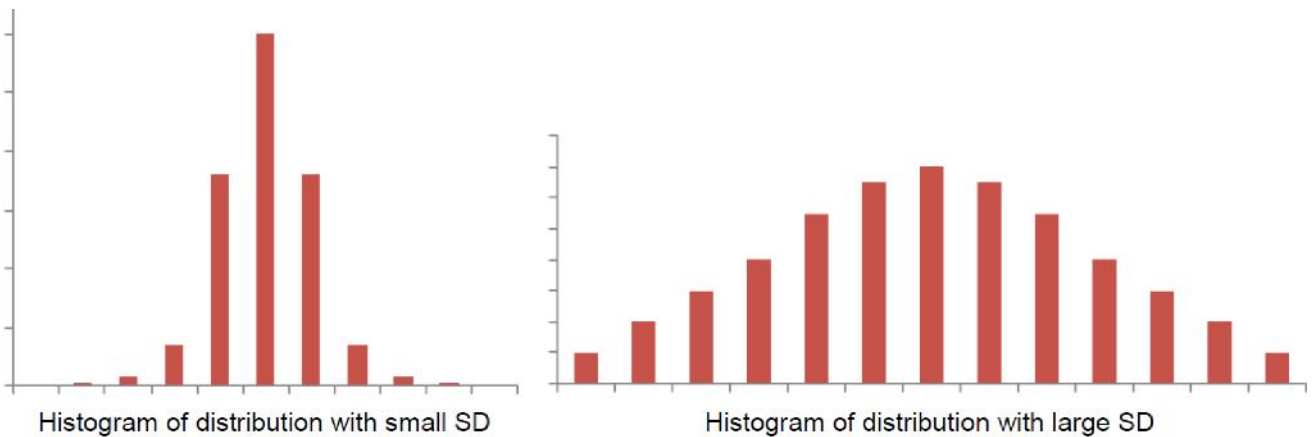


Figure 2: Histograms with response for a question with distribution

SD generally does not indicate "right or wrong" or "better or worse" -- a lower SD is not necessarily more desirable. It is used purely as a descriptive statistic. It describes the distribution in relation to the mean.

Standard Error

The Standard Error ("Std Err" or "SE"), is an indication of the reliability of the mean. A small SE is an indication that the sample mean is a more accurate reflection of the actual population mean. A larger sample size will normally result in a smaller SE (while SD is not directly affected by sample size).

Most survey research involves drawing a sample from a population. We then make inferences about the population from the results obtained from that sample.

Think about this. If the SD of this distribution helps us to understand how far a sample mean is from the true population mean, then we can use this to understand how accurate any individual sample mean is in relation to the true mean. That is the essence of the Standard Error. In actuality we have only drawn a single sample from our population, but we can use this result to provide an estimate of the reliability of our observed sample mean.

While the actual calculations for Standard Deviation and Standard Error look very similar, they represent two very different, but complementary, measures. SD tells us about the shape of our distribution, how close the individual data values are from the mean value. SE tells us how close our sample mean is to the true mean of the overall population. Together, they help to provide a more complete picture than the mean alone can tell us.

CASE STUDY 4: LAB SHIFT TABLES

A **shift table** is a **table** that displays the number of subjects who are low, normal or high at baseline and the **shift** at post-dose.

WHY DO WE NEED SHIFT TABLES?

The site that provides the laboratory results usually produces the normal ranges for numeric lab tests. The study team analyzes the lab results based on normal range and then flags or categorizes whether it is low, normal or high. In any clinical study it is crucial to look at the effect of the drug by comparing values at baseline and post baseline. Shift tables can be created for any data but it is very common to compare results from labs, vitals and ECG. It can be extremely useful in understanding the changes from one time point to another. They make it easier to figure out any abnormality in data because the layout helps to get the count of subjects at two time points together. The main purpose of shift tables in any clinical trial is to determine how the categorical result varies from baseline to post-dose.

In the given example the aim is to observe the change in the grades from baseline to post baseline in two different treatments. Consider a mock shell for the shift table to be produced for CTC grades at baseline to post-dose value for Fasting Glucose.

How do we understand/validate a shift table?

Fasting Glucose (mg/dL) Shift Table - Safety Population

Treatment	Baseline		Post-Baseline				
	Missing	Grade 1	Grade 2	Grade 3	Grade 4	Missing	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
2							
TRT A (N=19)	Missing	2 (5.9)	2 (50.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
1	Grade 1	13 (38.2)	0 (0.0)	3 (18.8)	5 (83.3)	4 (66.7)	1 (100.0)
	Grade 2	2 (5.9)	0 (0.0)	1 (6.3)	0 (0.0)	1 (16.7)	0 (0.0)
	Grade 3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 4	2 (5.9)	0 (0.0)	1 (6.3)	0 (0.0)	1 (16.7)	0 (0.0)
	Total	19 (55.9)	2 (50.0)	5 (31.3)	5 (83.3)	6 (100.0)	1 (100.0)
TRT B (N=15)	Missing	2 (5.9)	1 (25.0)	1 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)
3	Grade 1	12 (35.3)	1 (25.0)	9 (56.3)	1 (16.7)	0 (0.0)	0 (0.0)
	Grade 2	1 (2.9)	0 (0.0)	1 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 4	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Total	14 (44.1)	2 (50.0)	11 (68.8)	1 (16.7)	0 (0.0)	0 (0.0)
OVERALL (N=34)	Missing	4 (11.8)	3 (75.0)	1 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 1	25 (73.5)	1 (25.0)	12 (75.0)	6 (100.0)	4 (66.7)	1 (100.0)
	Grade 2	3 (8.8)	0 (0.0)	2 (12.5)	0 (0.0)	1 (16.7)	0 (0.0)
	Grade 3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Grade 4	2 (5.9)	0 (0.0)	1 (6.3)	0 (0.0)	1 (16.7)	0 (0.0)
	Total	34 (100.0)	4 (100.0)	16 (100.0)	6 (100.0)	6 (100.0)	1 (100.0)

Baseline is defined as the last non-missing value prior to first dose.
 Percentage is based on the Baseline total n

Figure 3: Sample Lab Shift table

Refer the numbering of the highlighted portions in the above table and read the corresponding explanation to see what these numbers signify.

Out of 19 subjects under Trt A, 13 subjects have grade 1 at baseline and three of them remain at grade1, five subjects shift to grade 2, four to grade 3 and one subject to the 4th grade post baseline. So it should be noted that the total number of subjects under post baseline grades should add up to number of subjects at baseline grade category in this row.

Total 19 should match the Total number of subjects under the respective treatment (N). The vertical total as well as horizontal total should get add up to this number.

As seen in Trt B, the total of 14 subjects at baseline does not seem to be consistent with the corresponding column total as well as the treatment total (N = 15). Focusing on the grade 1 category, it can be spotted that the counts from the post-baseline records do not sum up to 12. As a result the corresponding values from the Total row give the incorrect addition of 14, which in reality should have been 15. Consequently this has affected the counts in the 'Overall' group as well.

4. The second footnote of the table directs towards the denominator to be used for percentages. In this table, the percentages are calculated over the total of each grade at post-baseline. However the correct approach would be to use the totals at baseline value (point #5) for each of the corresponding three treatment groups. Thus the respective percentages would be 100% for these three totals. Many time the mock shell do not include the "Total" row which adds up the total of all grade counts. In such cases it is programmer's responsibility job to ensure that the total comes up to 100 before finalizing the table. Thus the shift tables, although look very simple with just some frequencies and percentages to be derived, need to be understood thoroughly and cross-checked very cautiously.

CASE STUDY 5: PLOTS: HOW THEY HELP US IN UNDERSTANDING DATA:

Box-and-Whiskers Plots

How understanding of Q1-Q3, Mean help me in understanding Box and Whisker plots?

To understand this lets consider an small data with values:77, 79, 80, 86, 87, 87, 94, 99 respectively.

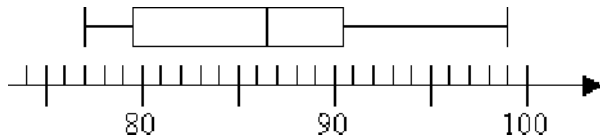
My first step is to find the median. Since there are eight data points, the median will be the average of the two middle values: $(86 + 87) \div 2 = 86.5 = Q2$

This splits the list into two halves: 77, 79, 80, 86 and 87, 87, 94, 99. Since the halves of the data set each contain an even number of values, the sub-medians will be the average of the middle two values.

$$Q1 = \frac{79 + 80}{2} = 79.5$$

$$Q3 = \frac{87 + 94}{2} = 90.5$$

The minimum value is 77 and the maximum value is 99, so I have: **min**: 77, Q1: 79.5, Q2: 86.5, Q3: 90.5, max: 99. Then my plot looks like this:



The top end of your box may also be called the "upper hinge"; the lower end may also be called the "lower hinge". The lower hinge is also called "the 25th percentile"; the median is "the 50th percentile"; the upper hinge is "the 75th percentile"

HOW TO IMPLEMENT IN SAS: Proc Boxplot:

This procedure produces side-by-side box and whisker plots for a continuous variable, displayed for each level of a categorical variable. The data set must first be sorted by the categorical variable. The syntax to produce a box plot is shown below. The plot statement first lists the continuous variable you wish to display, the second variable after the * is the categorical variable that will form the X-Axis categories.

```
proc sort data = sasdata2.business;
  by industry;
run;
```

```
proc boxplot data = sasdata2.business;
  plot sales * industry ;
run;
```

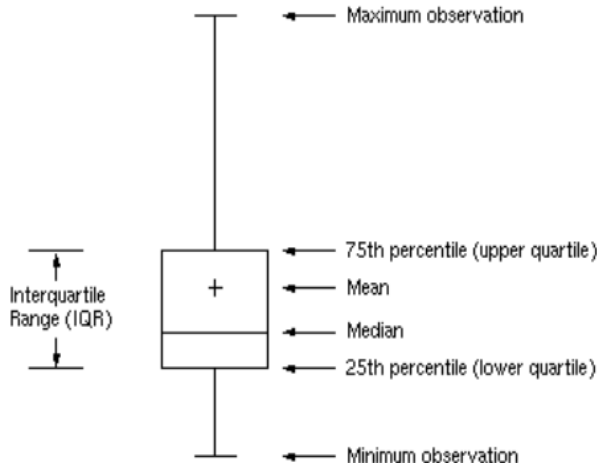



Figure 4: Anatomy of Box Plot

By default, the characteristics of the box plot are as:

- The length of the box represents the interquartile range (the distance between the 25th (lower quartile) and the 75th percentiles (upper quartile)).
- The plus inside the box represents the mean of the continuous variable.
- The horizontal line inside the box represents the median of the continuous variable.
- The vertical lines at the top and bottom of the box extend to the minimum and maximum values of the continuous variable.

You can change the display, so that SAS shows **outliers** in the graph, by using the **boxstyle=schematic** option.

```
proc boxplot data = sasdata2.business;  
  plot sales * industry /boxstyle=schematic.  
run;
```

What are scatter Plots and how does it help us:

The PROC for basic scatter plots is **PROC SGPLOT**.

What information does scatter plot provides me?

Scatter plots are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter plots show how much one variable is affected by another. The relationship between two variables is called their correlation. Scatter plots usually consist of a large body of data. The closer the data points come when plotted to making a straight line, the higher the correlation between the two variables, or the stronger the relationship.

```
proc sgplot data = testdata;  
  scatter x = Variable1 y = variable2;  
run;
```

If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a positive correlation. If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the

variables have a negative correlation. If the plot on the graph is scattered in such a way that it does not approximate a line (it does not appear to rise or fall), there is no correlation between the sets of data.

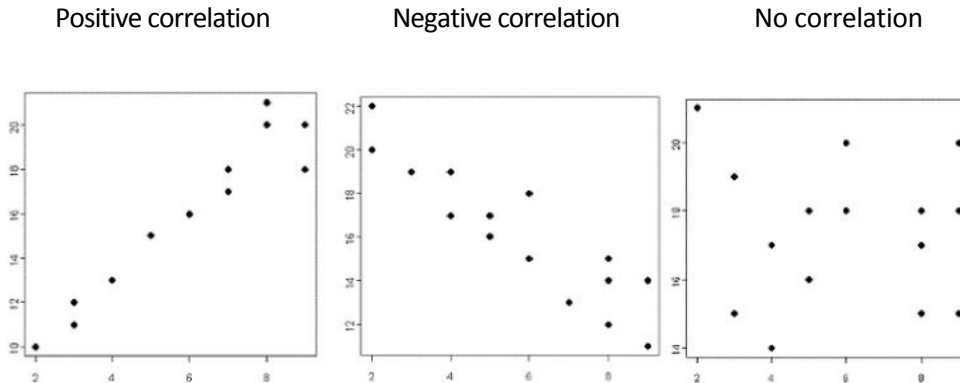


Figure 5: Showing 3 types of scatter plots with different correlation

CASE STUDY 6: KAPLAN MEIER GRAPH FOR SURVIVAL IN TWO TREATMENT GROUPS

Introduction:

In many clinical studies, either experimental or observational, the time until the occurrence of an event is the main outcome of interest; such an event can be either adverse, such as death or disease progression or recurrence of a tumor, or positive, such as birth or disease remission, or neutral, such as cessation of breast feeding or the completion of a task. In all these cases, it is conventional to talk about survival endpoints; the term survival may be misleading because the techniques are applicable to any well-defined event although traditionally death was the event of interest. Events in survival analysis are defined by a 'transition' from one discrete state to another at an instantaneous moment in time.

The typical feature of this kind of data is that not all the subjects will have experienced the event at the end of the follow-up period. This is called **censoring**, meaning that the observation period ended **without observing the event of interest, either because the subject has not yet experienced the event by** the end of follow-up period, or because the patient is lost to follow-up, or because the subject experiences a different event that makes further follow-up impossible. It is not possible in these cases to know if and when the subject would have experienced the event and we refer to this as **right censoring** meaning that a subject's follow-up stops before observing the event; while **left censoring** refers to when the event has occurred before the subject has entered the study.

PROC LIFETEST procedure can be used to create and provide tests to compare the survival curves for two different populations. The usual format of a SAS dataset for this analysis will comprise one observation per subject, a binary indicator variable (CENSOR) with a value of 1 indicating the time to the event of interest is complete or 0 indicating the time to the event was censored, a time to event (MONTHS), a treatment group (TRT) used to formulate a comparison and several covariates (SEX, AGE) which might also be considered to have an effect on survival.

```
PROC LIFETEST data=OncData;  
    TIME months*censor(0);  
    STRATA trtgrp;  
RUN;
```

Why Use a Kaplan-Meier Analysis?

To estimate a population survival curve from a sample

Consider the following Kaplan Meier Survival Probability graph for two treatment groups. Here the event of interest is death.

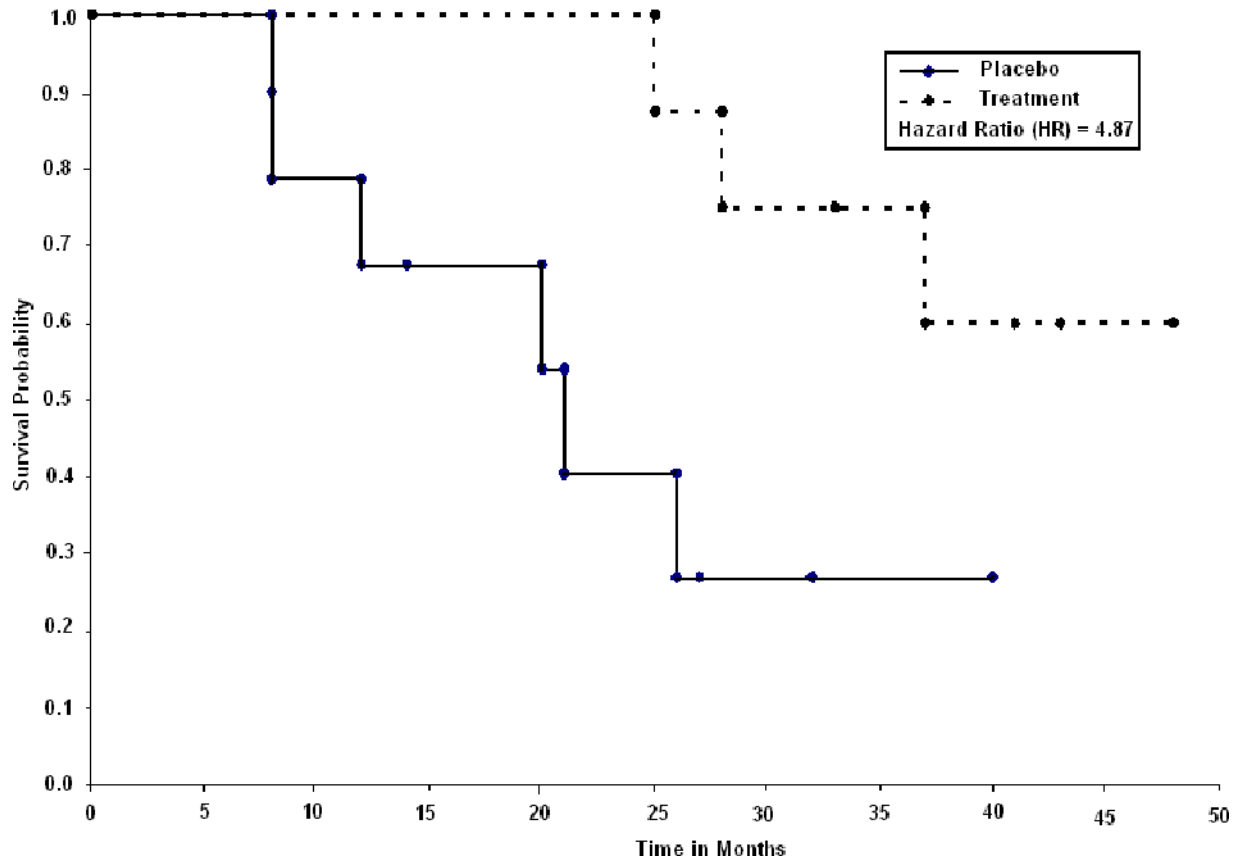


Figure 6: Sample Kaplan Meier Graph

Understanding the Graph:

- Above is an example Kaplan-Meier curve, with the x-axis representing time-to-event in months and y-axis representing the survival probability. For example: The first death in the treatment group occurs around 25th month, while the same for the placebo group happens at around the 8th month.
- The filled circles symbolize censored subjects i.e. these subjects cannot be evaluated any further for the event of interest. For example . Death cannot be recorded for subjects withdrawing from the study or lost to follow-up.
- Large vertical steps downward indicate a large number of deaths in the given time period, while large horizontal steps indicate few deaths have occurred during an interval. Thus the death incidence happens at a faster rate in the placebo group as compared to the treatment. Hence the possibility of survival in the latter is more as compared to that of the former

How do I check if graph is correct?

Now, let us see if the other results presented in the figure are consistent with the above three points. The Hazard ratio (HR) in a simple language is the ratio of the risk of experiencing the event of interest in the treatment population relative versus the control. It is already seen that the probability of event (i.e. death) in the placebo group is greater. Thus the HR should have been less than 1 due to the larger denominator, which is not consistent with the above result of HR = 4.87

So what could be the reason behind this mismatch? The possible answer to this is how the censoring variable has been defined and its value that has been entered in the SAS® procedure to calculate the HR. Suppose the censoring flag is defined as 1 when the record is censored and 0 when the death is observed. So while specifying in PHREG procedure which calculates the HR, the value 1 should be entered and not 0. This can be a quick check performed by the programmer before getting this as a QC finding from the validator.

Conclusion

We understand statistics is vast and at times complicated. We have tried to touch base on some basic topics in this paper to initiate a thought process to look at the numbers diligently and differently. One can be a successful programmer purely on the basis of the SAS expertise without any reference to Statistics. As we have seen from above scenarios that how does understanding of some basic concepts can help programmer to see beyond numbers and improve the quality of deliverable, and hence prevent re-work. If we can identify crucial data / programming issues ahead of time, we can improve efficiency of overall project. . The discrepancies in the data eventually affect the quality of the data and the date of the data base lock (DBL) .This in turn impacts the overall study timelines and consequently the budget of the trial. According to our experience we feel that a SAS programmer with some statistical knowledge is more efficient and has a competitive edge over a pure SAS programmer.

REFERENCES

- [1] <http://en.wikipedia.org/wiki/Standarddeviation>
- [2] <http://www.investopedia.com/terms/s/standard-error.asp>
- [3] <http://en.wikipedia.org/wiki/Biasofanestimator>
- [4] <https://www.lhup.edu/~dsimanek/scenario/errorman/distrib.htm>
- [5] Statistics - Third Edition, Freedman, Pisani & Purves
- [6] www.lexjansen.com/phuse/2013/is/IS05.pdf
- [7] SAS Programming in the Pharmaceutical Industry, Second Edition
- [8] <http://www.lexjansen.com/phuse/2013/is/IS05.pdf>
- [9] <https://communities.sas.com/t5/General-SAS-Programming/How-to-calculate-P-Value-in-SAS/td-p/181300>
- [10] <http://www.wuss.org/proceedings10/coders/29253COD-BECK.pdf>

ACKNOWLEDGEMENT

Authors would like to thank InVentiv health Clinical and our associate director John Durski for encouraging and supporting us to write and present papers.

CONTACT INFORMATION

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

Name: Arun Raj Vidhyadharan
Enterprise: inVentiv Health Clinical
Address: 500 Atrium Drive
City, State ZIP: Somerset, NJ 08873
Work Phone: 908.421.4667
E-mail: arunraj.vidhyadharan@inventivhealth.com

Name: Sunil Mohan Jairath
Enterprise: inVentiv Health Clinical
Address: 500 Atrium Drive
City, State ZIP: Somerset, NJ 08873
Work Phone: 862.240.4182
E-mail: sunil.jairath@inventivhealth.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.