

## Statistician in my Soul. Statistical Questions from SPA

Nadiia Hryhorenko, Experis Clinical, Kyiv, Ukraine

### ABSTRACT

Many times, statistical programmers cannot implement deep analyses of clinical data without advanced statistical knowledge, and answers to their questions may be too simple to bother the statisticians for help. In the article we address a few of the statistical issues programmers often face. Mastering these methods allows deeper understanding of the clinical data and may even help in detecting inconsistencies in the data analysis

### INTRODUCTION

Among numerous of problems which appear in the clinical trials data processing, we discuss three main issues in this article. The first issue is standard methods of analyzing non-normally distributed data. For example, the t-test is widely used to analyze results of clinical trials, and a normal data distribution is required for this test. General clinical data is normally distributed but it will be helpful to know what needs to be changed if we do not have expected distribution. In an attempt to find out if non-normal distributions could cause erroneous results, 10 papers were analyzed in the "Statistical Evaluation of Method-Comparison Data" article (Gaw Tzu Wu, Stanley L. Twomey, and Ralph E. Thlers, 1975). In three of them, data had non-normal distributions but standard methods were used. Moreover, two of these three non-normal distributions directly influenced the result. Since any errors in the results are unacceptable in clinical trials, we suggest a general approach to analyze both normally and non-normally distributed data. Additionally, in some cases a by-variable analysis (e.g., by visit) is needed and normality should be checked for every value of this variable. The suggested part of the program implements this analysis in SAS® and is suitable for any type of distribution.

Another issue reviewed in this paper is related to the Kaplan-Meier approach, which is often used in efficacy analyses. The Kaplan-Meier survival analysis indicates efficiency of the treatment and it involves estimations of time to the target event. The common issue that occurs here is why a value of 'NE' sometimes arises instead of numeric values for the median or quartiles. Typically, in ordinary statistics these values always exist, but it is not so with the Kaplan-Meier analyses. In this article we explain why quartiles may not exist and provide an algorithm for checking if the results are correct.

The next question is related to the confidence limits, namely which type of confidence limits is better to select: asymptotic or exact. Some procedures calculate both types by default and leaves the choice to a programmer. "Exact" sounds better, right? However, in some cases even exact method cannot be reliable enough, and in others, difference in results of asymptotic and exact methods is not significant whereas asymptotic is much more clear. We discuss different methods, their advantages and limitations, and give recommendations on using some of them.

### HOW TO CHOOSE A TEST

The purpose of this section is to provide an algorithm of analyzing data depending on its distribution. In efficacy analyses, we compare the effect of the investigated treatment between groups. Most often to find out if a difference exists, confidence intervals for a difference of means or a difference of proportions are calculated, which is equivalent to testing of hypothesis of means equality. We analyze some parameters for each group, e.g. results of laboratory tests. Therefore, we consider categorical variable (e.g. treatment group) as input variable and quantitative variable (e.g. a laboratory test result) as outcome variable. In addition, we assume that data are mutually independent and study design should guarantee it.

There are two types of tests: parametric and non-parametric. Parametric tests make an assumption about data distribution. This is often the assumption about normality of data distribution. The widely used t-test is parametric and it requires the data to be normally distributed. Non-parametric tests are also called "distribution-free", thus they can be applied both to normally and non-normally distributed data. Then you may ask why do we not use non-parametric tests in all cases? There are a couple of reasons and the

most significant one is that parametric tests have more statistical power than their non-parametric equivalents. In other words, they are more likely to detect a significant difference when it truly exists.

Let us formulate steps to provide the most suitable analysis assuming data independence:

1. Check data for distribution normality;
2. Depending on the first step result, select the best-fit test;
3. Realize this approach in SAS®.

The first step can be easily implemented using the CAPABILITY procedure. It performs the Shapiro-Wilk test which checks hypothesis on data distribution normality. As was mentioned in the introduction, we would like to produce a by-variable analysis. It means that normality should be checked for each value of by-variable. The following PROC CAPABILITY step performs appropriate analysis:

```
proc capability data = &dsin. normaltest outtable = _PROBN_;
  by &arm. &byvar.;
  var &var.;
run;
```

In the above code, the NORMALTEST option makes the PROC CAPABILITY perform the Shapiro-Wilk test. It tests null hypothesis that data sample came from normally distributed population. The OUTTABLE= option creates the data set \_PROBN\_ which contains all calculated statistics including p-value for normality test. Values of the variable \_PROBN\_ mean p-value for Shapiro-Wilk test. In case when the \_PROBN\_ variable is less or equal 0.05, we can reject null hypothesis and handle our data as non-normally distributed. In other cases, we can apply tests that require normal distribution of data.

To perform the second step, it is necessary to identify non-parametric tests equivalent to well-known parametric ones. We consider two cases: when input variable has two and more than two categories. In the Table 1 below the most common tests for each set of the variables are presented:

		Outcome variable	
		Normal	Non-normal
Input variable	2 values	Student's t-test	Wilcoxon-Mann-Whitney
	>2 values	Analysis of variance (ANOVA)	Kruskal-Wallis

**Table 1: Statistical hypothesis depending on number of input variable values and outcome variable distributio**

Note that this table is just a guide, and each real case should be considered separately.

Now let us take a look at the tests requirements and realization.

## THE STUDENT'S T-TEST

The Student's t-test tests null hypothesis that the means of two populations are equal. Its assumptions are as follows:

- Each of the two populations being compared should follow a normal distribution.
- The two populations being compared should have the same variance. If the sample sizes in the two groups being compared are equal, the Student's t-test is highly robust to the presence of unequal variances.
- The data used to carry out the test should be sampled independently from the two populations being compared. In general, this is not testable from the data, but if the data are known to be dependently sampled (that is, if they were sampled in clusters), then the classical t-tests discussed here may give

misleading results.

We assume that the PROC CAPABILITY has already checked normality of distribution.

The TTEST procedure envisages case of unequal variance. For this purpose the COCHRAN option might be added to the PROC TTEST statement in order to use the Cochran approximation. Anyway, the Satterthwaite approximation for unequal variances is included by default.

```
proc ttest data = &dsin.;  
  class &arm.;  
  var &var.;  
run;
```

The study design should guarantee independency of sampling. Population follows all sample constrains, i.e. if just women after 30 took part in the study then conclusions spreads just on women after 30.

## THE WILCOXON-MANN-WHITNEY TEST

The non-parametric equivalent for t-test is the Wilcoxon-Mann-Whitney (WMW) test. The WMW test checks null hypothesis about equality of distributions of both populations. These are following general assumptions for it:

- All the observations from both groups are independent of each other.
- The responses are ordinal (i.e., we can compare values and say which one of two observations is the greater)

As previously, we assume that independence is guaranteed by study protocol.

Since we consider the quantitative variable as outcome variable, the second assumption is also satisfied.

Note that this test is non-parametric, therefore the WMW test can be applied to arbitrary distributed data.

To realize the WMW test we can use the NPAR1WAY procedure with some options:

```
proc npar1way data = &an1 wilcoxon;  
  class &arm.;  
  var &var.;  
run;
```

In the PROC step above the WICOXON option is used to provide the WMW test for the analysis variable. In addition, this option includes the Kruskal–Wallis analysis by default.

The two previously discussed tests perform comparison of two populations. The following tests are applied for two or more population comparison.

## ONE-WAY ANALYSIS OF VARIANCE

One-way analysis of variance (ANOVA) tests the null hypothesis that two or more samples are drawn from populations with equal means. This technique can be used only for numerical outcome variable, and numerical or (usually) categorical input variable. ANOVA is “one-way” because we examine the influence of one input variable on one dependent variable. These are the following assumptions for this test:

- Outcome variable residuals are normally distributed (or approximately normally distributed).
- Variances of populations are equal.
- Outcomes for each group are independent and identically distributed normal random variables, i.e. each random variable has the same probability distribution as the others and all are mutually independent.

Several procedures are available to realize ANOVA in SAS®, and the ANOVA procedure is widely used for this purpose. The PROC ANOVA handles balanced data (that is, data with equal numbers of observations for every combination of the classification factors), whereas the procedures like the GLM procedure can analyze both balanced and unbalanced data. Because the PROC ANOVA takes into account the special structure of a balanced design, it is faster and it uses less storage than the PROC

GLM for the balanced data. The PROC NPAR1WAY also realizes a standard analysis of variance on the raw data. Which procedure is better to use should be decided for specific data. We suggest using the PROC NPAR1WAY because it is the most convenient for our purpose.

```
proc npar1way data = &anl anova;
  class &arm.;
  var &var.;
run;
```

In this code the ANOVA option produces one-way analysis of variance for the variable specified in the VAR statement.

## THE KRUSKAL–WALLIS TEST

The Kruskal–Wallis test by ranks is a non-parametric method for testing whether the sample medians are equal. It does not require equal sample sizes or normal distribution. However, there are the following simple assumptions:

- Outcome variable should be measured at the ordinal or continuous level.
- Input variable should consist of two or more categorical independent groups.
- Independence of observations, which means that there is no relationship between the observations in each group or between the groups.

This test is non-parametric, therefore the data can have non-normal distribution. However, if we can make the less stringent assumptions of an identically shaped and scaled distribution for all groups, except for any difference of medians, then the Kruskal-Wallis test compares medians for different groups. Otherwise, just mean ranks of the groups can be compared.

The PROC NPAR1WAY realizes the Kruskal-Wallis test the same as the WMW test:

```
proc npar1way data = &anl wilcoxon;
  class &arm.;
  var &var.;
run;
```

As we mentioned above the WOLCOXON option implements the Kruskal-Wallis test in case when the variable specified in the CLASS statement has more than two levels.

We have described variants of analyses depending on type of the data distribution. Therefore, now we can see that non-parametric tests also have requirements, but generally, they are easier to follow. It is necessary to remember that the incorrect test or non-fulfillment of the test requirements can lead to erroneous conclusion, which is unacceptable in clinical trial analyses. Next, we propose general part of code that implements analysis depending on normality in the first item in the Additional Materials.

## THE KAPLAN-MEIER APPROACH IN EFFICACY ANALYSIS

The Kaplan-Meier survival analysis, or product-limit estimator, is commonly used in processing of clinical trials results to handle time to event. E.g. in oncology studies it is often used to measure the fraction of patients living for a certain amount of time after treatment. The time starting from a defined point to the occurrence of event is called survival time. For subjects for whom the event does not occur before the end of the study, or was lost to follow up, or was drug withdrawn etc., the date of event is censored, i.e. is set to the date predefined by the protocol, and the survival time is calculated relatively to the censored date. For analysis, a set of data pairs for each subject that contains time to event (or time to censoring) and the subject's status (event occurrence or censored) is required. The survival analysis includes the plot with survival curve and several statistics. We are interested in the median and quartiles because they differ from corresponding sample statistics. In some cases the value of "NE" appears instead of numerical value of the median or quartile, whereas the sample median and quartiles always exist. Next, we formulate an analytical consideration for calculating target statistics and find out the reason of "NE" appearance.

Let  $t_1 < t_2 < \dots < t_D$  represent the distinct event times. For each  $i = 1, \dots, D$ , denote the number of

surviving units as  $n_i$ . It means that  $n_i$  subjects are at risk prior the time point  $t_i$ . Let  $d_i$  be the number of events at  $i$ th time point.

Then the product-limit estimator is given by the statement:

$$\hat{S}(t_i) = \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right)$$

This estimator is right continuous, i.e. events at  $i$ th time point are included in  $\hat{S}(t_i)$ .

The first quartile (or the 25th percentile) of the survival time is the time point beyond which 75% of the subjects in the population are expected to survive. Estimation of the first quartile is:

$$q_{0.25} = \min\{t_i | \hat{S}(t_i) < 0.75\}$$

If product-limit estimate is exactly equal to 0.75 at all interval from  $t_j$  to  $t_{j+1}$ , the first quartile is set to  $(t_j + t_{j+1})/2$ .

If  $\hat{S}(t_i) > 0.75$  for all values of  $i$  then the first quartile cannot be estimated. That is the reason why the first quartile does not exist. Next, we discuss it in more details below.

The general formula for estimating the  $100p$  percentile is:

$$q_p = \min\{t_i | \hat{S}(t_i) < 1 - p\}$$

Therefore, the median and the third quartile correspond to  $p = 0.5$  and  $p = 0.75$ .

With just the expression above we cannot predict non-existence of the quartile from the data. For this purpose, we formulate and prove the theorem about correlation between the number of events and existence of quartiles.

Theorem: if  $100p$  percentile does not exist then the total number of events is less than  $100p\%$  of the subject number in population. In the symbols for 25th percentile it looks as follows:

$$\prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right) > 0.75 \text{ for each } i = 1, \dots, D \Rightarrow \sum_{j=1}^D \frac{d_j}{N} < 0.25$$

To prove this equivalence we prove the equality:

$$\sum_{j=1}^D \frac{d_j}{N} = 1 - \prod_{j=1}^D \left(1 - \frac{d_j}{n_j}\right)$$

Let  $N$  to be the number of subjects in the population. Then the number of subjects live up to the first time point  $t_1$  equals to  $n_1 = N$ . For each  $i = 1, \dots, D$  number of subjects live up to the time point  $t_{i+1}$  equals to  $n_{i+1} = n_i - d_i$ .

The product-limit estimator for the time point  $t_i$  equals to:

$$\hat{S}(t_i) = \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right) = \prod_{j=1}^i \frac{n_j - d_j}{n_j} = \prod_{j=1}^i \frac{n_{j+1}}{n_j} = \frac{n_2}{n_1} \cdot \frac{n_3}{n_2} \dots \frac{n_{i+1}}{n_i} = \frac{n_{i+1}}{n_1} = \frac{n_{i+1}}{N}$$

The number of events that happen at the time point  $t_i$  equals to  $d_i = n_i - n_{i+1}$ . Consider sum in the left-hand-side of the target equation:

$$\begin{aligned} \sum_{j=1}^i \frac{d_j}{N} &= \frac{\sum_{j=1}^i d_j}{N} = \frac{d_1 + d_2 + \dots + d_i}{N} = \frac{n_1 - n_2 + n_2 - n_3 \dots + n_i - n_{i+1}}{N} = \frac{n_1 - n_{i+1}}{N} = \frac{N - n_{i+1}}{N} = 1 - \frac{n_{i+1}}{N} \\ &= 1 - \hat{S}(t_i) = 1 - \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right) \end{aligned}$$

Since this equation is satisfied for each  $i = 1, \dots, D$ , then:

$$\sum_{j=1}^D \frac{d_j}{N} = 1 - \prod_{j=1}^D \left(1 - \frac{d_j}{n_j}\right)$$

The theorem is proved.

The inverse of the abovementioned theorem is not generally true. The reason is in the data censoring. There are several types of censoring, but commonly used in clinical trials is right censoring. Right censoring means that a data point is above a certain value but it is unknown by how much. In practice, we select subjects for whom an event did not occur. The subject could be lost to follow-up, or discontinued from the study, or event did not occur to the end of the study. For them the date of the event is set to any date predefined by the protocol such as the date of the last visit or the date of any other event. Time to event is calculated from the date of start to follow-up to the date of event when it occurred or the censored date when it did not. There is a question if censored dates influence the product-limit estimation. Of course, they do, but not in a way we expected.

Formally, a time point, which corresponds to the censored date, provides 1 as multiplier to the product-limit estimator ( $1 - \frac{d_j}{n_j} = 1 - \frac{0}{n_j} = 1$ ).  $\hat{S}(t)$  is not calculated at these points at all, whereas number of the censored dates affects the estimation for the next point. An event does not occur at the censored point, but the patient is still not included in the estimation for the next point. Let us remind some designations: we consider  $N$  subjects,  $n_i$  equals to the number of subjects alive just prior the time point  $t_i$ , and  $d_i$  equals to the number of events at  $i$ th time point. In the Table 2 below there is an example illustrated how censoring influences the median existing:

Time point * - censored	$n_i$	$d_i$	$\left(1 - \frac{d_i}{n_i}\right)$	$\prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right)$	
12	10	1	0.9	0.9	
16*	9	0	1		
21	8	1	0.875	0.7875	
25*	7	0	1		
26	6	1	0.833	0.656	- 25th percentile
29*	5	0	1		
30	4	1	0.75	0.492	- 50th percentile
32*	3	0	1		
33*	2	0	1		
35*	1	0	1		

**Table 2: An example illustrates how censored values influence the product-limit estimation**

In this example, the number of subjects equals to 10, the number of events – 4. The percentage of events  $100 \cdot \frac{4}{10} = 40\%$  which is less than 50, but the 50th percentile exists and equals to 30. This happens because censored values are not concentrated at the end, but mixed up with occurred events.

Next, we consider an example where we can apply the theorem.

Assume that we investigate an effect of the treatment in two variants compared with placebo. For this purpose, 10 patients were randomized to each treatment group. We calculate time from the randomization to the target event. In case when this event does not occur during the study, we use cut-off date as censored date of this event. The results of life-test are presented in the Display 1. The first item in the Additional Materials generates data set and provides following output.

	Active (N=10)	Active+ (N=10)	Placebo (N=10)
Patients with event	4 (40.0%)	6 (60.0%)	8 (80.0%)
Patients without event	6 (60.0%)	4 (40.0%)	2 (20.0%)
Median	NE	20.5	15.7
Q1, Q3	20.6, NE	15.7, NE	11.4, 17.8

### Display 1: Output for Time Point Analysis

The target event occurred for 4 patients in the Active treatment group, 6 patients in the Active+ treatment group, and 8 patients among Placebo treated subjects. Since by default percentage for events is displayed, we can just look at it and make a conclusion. For the Active treatment 40% of the patients have the event, other 60% were censored. The median and the third quartile do not exist, and we know this result is correct, because  $40\% < 50\%$  and  $40\% < 75\%$ . Next, 60% of the Active+ group have the event and 40% were censored. For this group just the third quartile does not exist, therefore  $60\% < 75\%$ . In addition, for the Placebo treatment all quartiles exist, and the percentage of the patients with occurred event equals to  $80\% > 75\%$ .

As we discussed above the quartiles could exist when the percentage is small, but anyway, if the quartile does not exist, now we understand what the reason is.

### CONFIDENCE LIMITS: ASYMPTOTIC VS EXACT

Let us imagine that we run a procedure to calculate the confidence interval for the proportion. For instance:

```
proc freq data = anl2 noprint;
  by TRTGROUP;
  tables RESP / binomial (level = "Y") out = freqs;
  output out = datain BINOMIAL;
run;
```

Note that the FREQ procedure with the BINOMIAL option in the TABLES statement requires work around to have both levels of analysis variable for every set of variables specified in the BY statement. More details about test data for this example are in the second item in the Additional Materials section.

In the result of performing this PROC step we get the result data set DATAIN which is specified in the OUTPUT statement. Table 3 presents this data set.

Obs	Treatment Group	Number of Subjects	Binomial Proportion P	Lower CL, Binomial Proportion	Upper CL, Binomial Proportion	Exact Lower CL, Binomial Proportion	Exact Upper CL, Binomial Proportion
1	Active	10	0.6	0.29636	0.90364	0.26238	0.87845
2	Placebo	10	0.2	0.00000	0.44792	0.02521	0.55610

**Table 3: Result data set DATAIN of the PROC FREQ procedure with the BINOMIAL option**

By default, we get an output that contains both asymptotic and exact confidence intervals. Which one should we select, and what is the difference between them?

The standard Wald asymptotic confidence limits are based on the normal approximation to the binomial distribution. The formula for Wald confidence limits is clear:

$$\hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where  $\hat{p} = n_s/n$  is the sample proportion of successes, and  $z$  is the  $100(1 - \frac{\alpha}{2})$  percentile of the standard normal distribution.

However, this approximation is unreliable for small samples and in case when the success probability is close to 0 or 1, because it is based on central limit theorem.

The Clopper-Pearson interval is based on the cumulative probabilities of the binomial distribution, and does not approximate distribution, but exactly corrects it. That is why the other name of the Clopper-Pearson interval is “exact”. The formula for confidence limits is more complicated:

$$p_L = \left( 1 + \frac{n - n_S + 1}{n_S F\left[\frac{\alpha}{2}; 2n_S, 2(n - n_S + 1)\right]} \right)^{-1}$$

$$p_U = \left( 1 + \frac{n - n_S}{(n_S + 1) F\left[1 - \frac{\alpha}{2}; 2(n_S + 1), 2(n - n_S)\right]} \right)^{-1}$$

Where  $F[\alpha; n_1, n_2]$  is the  $\alpha$ th percentile of the  $F$  distribution with  $n_1$  and  $n_2$  degrees of freedom.

For the Clopper-Pearson interval for any fixed parameter value, the actual coverage probability can be much larger than the nominal confidence level unless  $n$  is quite large. For instance, the true coverage rate of a 95% Clopper–Pearson interval may be well above 95%. Mostly, it is out of our interests to have the confidence interval wider than the nominal coverage, thus another method is preferable. For big  $n$  even Wald interval is better to use.

A compromise solution is the confidence interval based on inverting the approximately normal test that uses the null, rather than the estimated, standard error. It is the Wilson score interval with the formula for the confidence limits:

$$\frac{\hat{p} + \frac{z^2}{2n}}{1 + \frac{z^2}{n}} \pm \frac{z}{1 + \frac{z^2}{n}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} + \frac{z^2}{4n^2}}$$

Where  $z$  is the  $100(1 - \frac{\alpha}{2})$  percentile of the standard normal distribution.

The Wilson score interval is an improvement over the normal approximation interval in which the actual coverage probability is closer to the nominal value. This interval has good properties even for a small number of trials and an extreme probability. Moreover, it provides the shorter interval with actual coverage probability less than but very close to nominal level.

There are several more methods for calculating confidence limits for proportion, such as Adjusted Wild interval (Agresti-Coull), Jeffreys interval, Arcsine transformation and so on. Which one should be selected depends on data, but we can make at least some general recommendation.

The Wald interval is reliable when  $n$  is big and  $\hat{p} \approx \frac{1}{2}$ , and it is the clearest and fastest method to calculate.

The Clopper-Pearson interval is more reliable for small  $n$  and small  $\hat{p}$ , but it provides interval with the actual coverage probability larger than the nominal.

The Wilson interval has a much better performance than the Wald interval and the Clopper-Pearson interval, the only point is that the actual coverage probability could be less than the nominal level in case of small  $n$ .

## CONCLUSION

We have discussed just a small portion of the issues, which we meet in our work. We hope that our recommendations will help to prevent mistakes related to chosen statistical methods. Note that the theory we proposed is just a guide, and every set of data should have its own approach. Unified approach is not possible, because every study is unique and requires special analyses. It is statisticians' area of work, but we should understand methods and results. In addition, there are cases, which were not covered in this article such as other types of input or outcome variable, or other methods for confidence limits



computation. These and other issues require additional investigation in the future.

## ADDITIONAL MATERIALS

1. The following part of program implements comparing of two groups by calculating p-value for the null hypothesis about equality of means or medians depending on the distribution. To be able to use this code you have to predefine macro variables: MAXVIS = <number of the last analysis visit>, ANL = <name of the analysis data set>, ARM = <name of the variable which contains treatment groups identifiers>, VAR = <name of the analysis variable>.

```

/*create empty data set to save calculated statistics there*/
data result;
run;
/*create macro to call statistical procedures*/
%macro calc_pvalue;
/*cycle by visit from 1 to maximum number of visit*/
%DO visit = 1 %TO &maxvis.;
/*check the data at visit for distribution normality*/
data datain;
    set _PROBN_ (where = (VISITNUM = &visit));
    call symput ('pvalue', _PROBN_);
run;
/*use the test for non-normal distribution*/
%IF &pvalue. LE 0.05 %THEN %DO;
    proc nparlway data = &anl (where = (VISITNUM = &visit)) wilcoxon;
        class &arm.;
        var &var.;
        output out = statdiff wilcoxon;
    run;
    data statdiff;
        set statdiff;
        VARIABLE = "&var.";
        VISITNUM = &visit.;
        METHOD = "Wicoxon";
        PVALUE = P2_WIL;
    run;
%END;
/*use the test for normal distribution*/
%ELSE %DO;
    proc ttest data = &anl. (where = (VISITNUM = &visit));
        class &arm.;
        var &var.;
        ods output TTests = statdiff ;
    run;
    data statdiff;
        /*keep results assuming equality of variances; otherwise
        METHOD = "Satterthwaite" can be selected*/
        set statdiff (where = (METHOD = "Pooled"));
        VARIABLE = "&var.";
        VISITNUM = &visit.;
        METHOD = "t-test";
        PVALUE = PROBT;
    run;
%END;
/*keep results of calculations for every visit in one data set*/
data result;
    set result statdiff;

```

```

        if not missing(PVALUE);
        keep VARIABLE VISITNUM METHOD PVALUE;
    run;
%END;
%mend calc_pvalue;
/*call macro*/
%calc_pvalue;

```

2. Test data for analysis are presented in the Table 4Table 4. The following program performs survival analysis and produces an output with statistics.

Obs	USUBJID	TRTGROUP	RANDDT	EXACERB	CUTOFF
1	TEST001	Active	1Jan2015	.	10Mar2018
2	TEST002	Active	12Feb2015	27Oct2016	10Mar2018
3	TEST003	Active	10Jan2015	12Oct2016	10Mar2018
4	TEST004	Active	3Sep2015	.	10Mar2018
5	TEST005	Active	20Jun2015	3Dec2016	10Mar2018
6	TEST006	Active	3Mar2016	.	10Mar2018
7	TEST007	Active	3Mar2015	11Feb2016	10Mar2018
8	TEST008	Active	11Jan2016	.	10Mar2018
9	TEST009	Active	27Apr2016	.	10Mar2018
10	TEST010	Active	13Aug2015	.	10Mar2018
11	TEST011	Active+	17Apr2015	.	10Mar2018
12	TEST012	Active+	9Apr2016	20Dec2017	10Mar2018
13	TEST013	Active+	27Aug2015	1Jan2017	10Mar2018
14	TEST014	Active+	11Jan2016	13Jan2018	10Mar2018
15	TEST015	Active+	25Mar2015	28Feb2016	10Mar2018
16	TEST016	Active+	12Oct2015	.	10Mar2018
17	TEST017	Active+	1Sep2015	30May2016	10Mar2018
18	TEST018	Active+	23May2015	.	10Mar2018
19	TEST019	Active+	30Jun2016	.	10Mar2018
20	TEST020	Active+	29Feb2016	18Jun2017	10Mar2018
21	TEST021	Placebo	1Jan2016	11Apr2017	10Mar2018
22	TEST022	Placebo	13Apr2016	.	10Mar2018
23	TEST023	Placebo	13May2015	30Oct2016	10Mar2018
24	TEST024	Placebo	12Dec2015	11Jun2016	10Mar2018
25	TEST025	Placebo	22Mar2016	3Nov2016	10Mar2018
26	TEST026	Placebo	10Sep2015	2Jan2017	10Mar2018
27	TEST027	Placebo	18Jun2015	.	10Mar2018
28	TEST028	Placebo	7Feb2015	16Feb2016	10Mar2018
29	TEST029	Placebo	29Apr2015	8Apr2016	10Mar2018
30	TEST030	Placebo	16May2015	25Oct2016	10Mar2018

**Table 4: Data set for survival analysis**

```

/*generate variable for the time to event and censoring flag*/
data anl;
    set asl;
    if missing(EXACERB) then do;
        EVENTDT = CUTOFF;
        CENSFL = 0;
    end;
    else do;
        EVENTDT = EXACERB;
        CENSFL = 1;
    end;

```

```
end;
TIMETEM = (EVENTDT - RANDDT + 1) / 30.25;
run;
proc sort data = anl;
  by TRTGROUP;
run;
/*calculate number of subjects with and without event*/
proc freq data = anl;
  table CENSFL / out = freqs;
  by TRTGROUP;
run;
/*perform survival analysis*/
proc lifetest data = anl method = km;
  time TIMETEM * CENSFL (0) ;
  by TRTGROUP ;
  ods output Quartiles = surv;
run;
```

3. The following part of the program convert our test data to the needed format and call the PROC FREQ to calculate binomial proportions and confidence limits for them.

```
/*create analysis data set for the calculating of confidence limits for
proportion*/
data anl2;
  set anl (where = (TRTGROUP in ("Active", "Placebo")));
  if CENSFL = 0 then RESP = "Y";
  else RESP = "N";
  label TRTGROUP = "Treatment Group";
run;
/*calculate confidence limits for proportion*/
proc freq data = anl2 noprint;
  by TRTGROUP;
  /*save frequencies to the data set FREQS*/
  tables RESP / binomial (level = "Y") out = freqs;
  /*save binomial statistics to the data set DATAIN*/
  output out = datain BINOMIAL;
run;
```

## REFERENCES

1. Lawrence D. Brown, T. Tony Cai, Anirban DasGupta. 2001. "Interval Estimation for a Binomial Proportion." *Statistical Science*, 16, 2:101-117
2. Agresti A., Brent A. Coull.1998. "Approximate Is Better than "Exact" for Interval Estimation of Binomial
3. Proportions". *The American Statistician*, 52, 2:119-126.
4. Zhang, D. 2005. Lecture Notes (Modified from Dr. A. Tsiatis' Lecture Notes). ST 745 Analysis of Survival Data, 11-14.

## RECOMMENDED READING

- SAS® Procedures Documentation. Available at <http://support.sas.com/documentation/>
- Annual Review Public Health. "Censoring issues in survival analysis". Available at <https://www.annualreviews.org/doi/pdf/10.1146/annurev.publhealth.18.1.83>

## CONTACT INFORMATION

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

Nadiia Hryhorenko  
Experis Clinical / Intego Group, LLC  
+380 (44) 5007020 (Ext. 2434)  
nadiia.hryhorenko@intego-group.com  
[www.intego-group.com](http://www.intego-group.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.