Study Design with SAS®: Estimating Power with Monte Carlo Methods
Robert W. Graebner, Pharmaceutical Research Associates, Inc., Lenexa, KS

ABSTRACT

An important aspect of study design is the determination of the sample size necessary to ensure that the study objectives are likely to be met. In addition to the total number of samples, the spatial and temporal distribution of sampling effort can have a significant impact on the power of the study. Estimating the power of a test is not always a simple task and it is often desirable to estimate power for a range study designs. The use of Monte Carlo methods allows a variety of designs to be simulated and the results compared to find the design that meets the desired objectives most efficiently.

The methods presented in this paper give examples of how SAS can be used to estimate the power of a study design by repeatedly simulating the design with randomly generated data. While this general method can be used with a variety of tests, this paper will provide examples of how to apply this procedure to a test for trend.

INTRODUCTION

When designing a monitoring study there are many design decisions to be made. It is possible to vary the length of the study, the frequency of sample collection and the number of samples to be collected at each event. The optimal distribution of sampling effort will depend on the minimum trend that must be detected in relation to the spatial and temporal variation within the study population.

One possible study design goal would be to minimize cost for a specified minimum detectable trend, power and significance level. If the study period is fixed, then it remains to be determined which combination of sample size and sampling frequency will result in the least effort or cost. To accomplish this requires comparing alternative designs, which can be very time consuming.

When conducting a test to determine if a positive trend is present, the null hypothesis (H_0) states that the trend is less than or equal to zero. The alternate hypothesis (H_A) states that the trend is greater than zero. By generating random sample data from a distribution that can be assumed to be similar to the study population, it is possible to simulate the results of the test under conditions where the null hypothesis is in fact false. By repeating this process numerous times, it is possible to estimate the power of the study. By running simulations of variations of the study design it is possible to gain insight into the relationship between power and sample size and sampling frequency. This information can then be used to arrive at the most efficient design.

METHOD

The Monte Carlo method of power estimation involves generating a series of random samples from a distribution that represents the study population of interest. The data generated for each sample are then modified in such a way that the null hypothesis is false. This can be accomplished by such methods as adding a known difference to all observations in one sample or by adding a known trend to samples that represent a time series. The test statistic is then calculated and compared to the critical value for the desired level of significance. Because the generated data

represent a case in which the null hypothesis is in fact false, the percentage of simulations that result in the rejection of the null hypothesis can be used as an estimate of the power of the test under consideration.

The method of power estimation presented here consists of the following steps:

1. Remove any trend from the historical data.
2. Estimate distribution parameters using the resulting data.
3. Generate a set of random observations for each sampling event and add a trend of a known magnitude.
4. Calculate test statistic and compare to critical value.
5. If H_0 is rejected, increment the number of successes by one.
6. Repeat steps 1 through 5.
7. Calculate estimate of power as the percentage of trials that were successful in rejecting H_0.
8. Repeat for different sample sizes, program lengths or trends to obtain power curves.

Removing any Trend from Historical Data

If the data to be used to estimate the distribution parameters were obtained at more than one point in time, a trend may be present that will distort the parameter estimates. One method of removing a simple linear trend from a set of existing data is to use linear regression. A simple linear trend model of the form

$$y_t = \beta_0 + \beta_1 t + e_t$$

can be fitted to the data and a detrended set of data can be obtained by

$$d_t = y_t - \beta_0 - \beta_1 t$$

which is equal to e_t, the residual from the regression. Using PROC REG it is simple to fit such a model and output the residuals to a data set. In the example below, data from several years were used and a separate model was fitted to the data from each calendar month since it was not assumed that the trends were same for each month. Figure 1 shows an example of historical data with the fitted regression line.

```
proc reg;
   by month;
   model conc=date;
   output out=dtrend r=concres;
run;
```
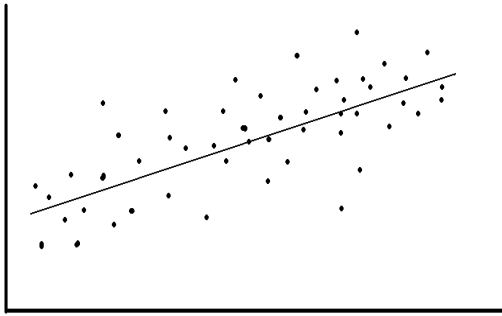
Figure 1: Historical data with fitted regression line.

Estimating Distribution Parameters Using Resulting Data

PROC MEANS can be used to obtain estimates of the normal distribution parameters for each month using the detrended data. The parameter estimates are output to a SAS data set for use in generating random observations from this distribution. Figure 2 shows the relationship between the residuals and the parameter estimates.

```
proc means;
   var concres;
   by month;
   output out=distparm mean=meanx std=stdevx;
run;
```
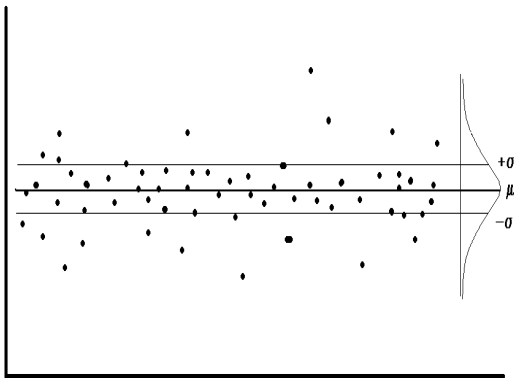

Figure 2: Data with trend removed and parameter estimates.

Generation of Random Observations with Known Trend

The RANNOR() function generates random values from a normal distribution with mean 0 and standard deviation of 1. By multiplying the resulting value by the standard deviation of the study population and adding the mean, it is possible to generate random values from the same distribution the study population has.

In the example below, 10 random values are generated to simulate a sample size of 10 at each sampling event. From theses values a sample mean is calculated and a trend of a known magnitude is added.

```
%let trend = 0.5;

data sampres;
   set distparm;
   type = 'ANNUAL';
   array avex{0:9};
   do iter=1 to 5000;
      do year=0 to 9;
         sumx=0;
         do n=1 to sampsize;
            sumx = sumx+meanx+stdevx*rannor(0);
         end;
         avex{I} = (sumx/sampsize) + (year*&trend);
      end;
      output;
   end;
run;
```

Calculate Test Statistic

For each simulation, the test statistic is calculated using standard SAS procedures or a SAS macro. The resulting test statistic is then compared to the critical value based on the specified significance level. If the null hypothesis is rejected, then the number of successes is incremented by one. Because a trend of the specified magnitude is present in the generated data, the percent of trials that result in the rejection of the null hypothesis can be used as an estimate of statistical power. The power is then output into a data set along with the value of the parameters that were varied across simulations. This can include sample size at each sample event, sampling frequency, number of sampling events and minimum detectable trend.

```
data testres;
   set sampres;
   %seakend(avex1,avex2,avex3,avex4,avex5
            avex6,avex7,avex8,avex9,avex10,res)
run;

proc sort data=testres;
   by type trend;
run;

proc freq data=testres;
   tables res / out = simres;
   by type trend;
run;

data powerres;
   set simres;
   keep type trend percent;
   rename percent=power;
   if res = 1;
run;
```

Plotting Power Curves

PROC PLOT or PROC GPLOT can be used to create power curves from the observations in the results data set. Producing

graphs with power curves for various study design alternatives can facilitate quick comparison. Such graphs are also useful in study design proposals to justify design decisions. In Figure 3, quarterly and annual sampling programs are compared for a 10-year monitoring program. Assume the study design criteria state that an annual trend of 90 or greater must be detected with 80 percent power. Figure 3 can be used to show that a quarterly sampling program will be adequate while an annual program will not.

```
proc gplot;
  plot power*trend=type;
  symbol1 i=spline l=2O  c=black;
  symbol2 i=spline l=1  c=black;
  label type='SAMPLING FREQUENCY';
  title  'Power to Detect Trend vs Trend';
  title2 'Quarterly and Annual Sampling';

run;
```
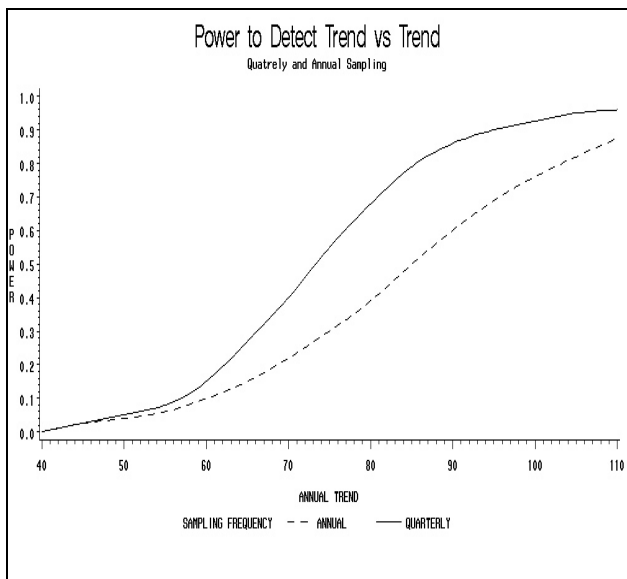

Figure 3: Power curves.

CONCLUSION

The programming techniques presented in this paper can be used to obtain estimates of statistical power for a wide range of studies and a wide range of test statistics. By creating a SAS program to perform the simulation, it is easy to repeat the simulation using a variety of study parameters in order to gain insight into the interrelationship of the study parameters and the power of the study.

REFERENCES

Noreen, E. W. (1989) Computer Intensive Methods for Testing Hypotheses, Wiley, New York.

Shumway, R. H. (1988) Applied Statistical Time Series Analysis, Prentice-Hall, New Jersey.

ACKNOWLEDGMENTS

CONTACTING THE AUTHOR

Robert Graebner
PRA, Inc.
16400 College Blvd.
Lenexa, KS 66219

graebnerbob@pra-ww.com
graetech@kc.net