

# ASP Trend Analysis Guidance

## Purpose

The purpose of the Accident Sequence Precursor (ASP) R-script is to compute trend lines and confidence intervals for the ASP data (both precursor counts and the integrated ASP index). The trend lines and confidence intervals are calculated to support the annual ASP Program report.

## Data

The ASP data consist of precursor counts, the integrated ASP index, and exposure times for each year over an extended period. Each precursor is assigned to more than one precursor category based on reactor technology, event type, and risk significance. There are eight of these overlapping categories and one combined category that counts each precursor once. Therefore, the R-script performs an analysis for nine different, but not mutually exclusive, precursor categories.

The ASP data contain counts for all nine precursor categories, the integrated ASP index, and exposure times for the past 20 years (the current data set contains values from 1999–2019). The ASP data also contain counts for the combined category and exposure times for a much longer period (the current data set contains these values from 1969–2019).

## Method

Since the data are different (discrete values for precursor counts and continuous values for the integrated ASP index), different techniques must be used to compute trend lines and confidence intervals. The R-script is designed to perform separate analyses for the precursor counts and the integrated ASP index.

The R-script uses the same time frames for each analysis. The primary trend lines and confidence intervals are calculated for the last 10 years, but additional trend lines and confidence intervals are calculated for the last 20 years as well as the entire set of historical data if data exist for more than 20 years (e.g., the combined precursor category).

## Precursor Counts

The R-script performs an identical analysis for each combination of precursor category and time period. For each combination, the R-script calculates trend lines and confidence intervals based on a Poisson regression model. In this model, the natural logarithm of the rate of occurrence is assumed to be a linear function of time. Mathematically, this is specified as:

$$\ln(\lambda) = \alpha + \beta t$$

This model is described in Section 7.2.4 of [NUREG/CR-6823](#), “Handbook of Parameter Estimation for Probabilistic Risk Assessment,” (ADAMS Accession No. ML032900131).

The R-script uses numerical techniques (iteratively reweighted least squares) to estimate values for the two parameters, the normal distribution to calculate pointwise confidence intervals for the

rate of occurrence at a single fixed time, and the chi-square distribution to calculate simultaneous confidence bands for the rate of occurrence at all times.

### Integrated ASP Index

For each time period, the R-script calculates trend lines and confidence intervals based on a log-linear regression model. In this model, the natural logarithm of the integrated ASP index is assumed to be a linear function of time. Mathematically, this is specified as:

$$\ln(y) = \alpha + \beta t$$

This model is mentioned, but not described in detail in [NUREG/CR-6823](#). Additional information and results for this simple linear regression model can be found in many elementary statistics textbooks.

The R-script uses analytical formulas to estimate values for the two parameters, the Student's t distribution to calculate pointwise confidence intervals for the integrated ASP index at a single fixed time, and the Working-Hotelling procedure to calculate simultaneous confidence bands for the integrated ASP index at all times.

### **Results**

The R-script produces a comma separated value (.csv) file for each of the analyses performed. [Table 1](#) describes the twelve variables generated from the analysis of precursor counts. [Table 2](#) describes the four variables generated from the analysis of the integrated ASP index. In addition, the R-script produces another .csv file that contains p-values associated with each regression model.

In statistics, p-values arise in the context of hypothesis testing. For a trend analysis, the p-value characterizes the level of statistical significance for the regression model. Traditionally, a p-value less than 0.05 indicates that a statistically significant trend exists. If a p-value is less than 0.05, the user rejects the null hypothesis that no trend exists. Conversely, if a p-value is greater than 0.05, the user fails to reject the null hypothesis that no trend exists.

### **Program Execution**

The R-script requires two additional files to run. The first file is another R-script containing the functions used to calculate the trend lines and confidence intervals. The second file is the .csv file containing the ASP data.

The R-script is designed for maximum flexibility. Before executing the script, the user specifies three directories. The first directory is the location where the additional R-script file is contained, the second directory is the location where the ASP data is contained, and the third directory is the location where the output files will be located. The R-script creates this directory if it does not already exist.

After defining the three directories, the user specifies the names of the additional R-script and the ASP data file. The default names are "ASP Functions.R" and "ASP Data.csv," respectively.

Finally, the user can specify the confidence level for the confidence intervals. The default value is 0.90. The R-script will use the same confidence level for all confidence interval calculations.

To execute the R-script using R Studio, press the button marked "Source." This button is in the top right corner of the Source pane. The Source pane is in the top left corner of R Studio.

### **Excel Spreadsheet**

To understand and validate the results of the R-script, an Excel spreadsheet was developed to perform the Poisson regression and the log-linear regression. This spreadsheet can be used, if desired, to perform the calculations performed by the R-script.

### **References for Calculations**

1. U.S. Nuclear Regulatory Commission, "Applying Statistics," NUREG-1475 (Revision 1), March 2011, ADAMS Accession No. [ML11102A076](#).
2. U.S. Nuclear Regulatory Commission, "Handbook of Parameter Estimation for Probabilistic Risk Assessment," NUREG/CR-6823, September 2003, ADAMS Accession No. [ML032900131](#).
3. Idaho National Engineering Laboratory, "Events in Time: Basic Analysis of Poisson Data," [EGG-RAAM-11088](#), September 1994.

**Table 1. Variables Generated from Each Trend Analysis of the Precursor Counts**

<b>Name</b>	<b>Description</b>
Year	Calendar year
x	Observed precursor count
T	Observed exposure time
MLELow	Lower bound of the pointwise confidence interval for the rate of occurrence
MLE	Maximum likelihood estimate for the rate of occurrence
MLEHigh	Upper bound of the pointwise confidence interval for the rate of occurrence
TrendLow	Lower bound of the simultaneous confidence band for the rate of occurrence
Trend	Predicted rate of occurrence from the trend line
TrendHigh	Upper bound of the simultaneous confidence band for the rate of occurrence
PredEventLow	Lower bound of the simultaneous confidence band for the number of occurrences
PredEvent	Predicted number of occurrences from the trend line
PredEventHigh	Upper bound of the simultaneous confidence band for the number of occurrences

**Table 2. Variables Generated from Each Trend Analysis of the Integrated ASP Index**

<b>Name</b>	<b>Description</b>
Year	Calendar year
Lower	Lower bound of the simultaneous confidence band for the integrated ASP index
Fit	Predicted integrated ASP index from the trend line
Upper	Upper bound of the simultaneous confidence band for the integrated ASP index