TECHNICAL REPORT

Data and Statistics 101: Key Concepts in the Collection, Analysis, and Application of Child Welfare Data

CONTENTS

August 20, 2015

Philip Osteen, Ph.D., Assistant Professor
Florida State University College of Social Work
296 Champions Way
Tallahassee, Florida 32306
posteen@fsu.edu
850.644.4751

KEY WORDS
child welfare statistics, child welfare data, data analysis

Funded through a contract with the Florida Institute for Child Welfare

FLORIDA INSTITUTE for CHILD WELFARE

Executive Summary

This technical report provides an overview of key concepts involved in data analysis and includes demonstrations of existing community-based care data on the application of statistical techniques for data analysis.

The Florida Department of Children and Families Office of Child Welfare (DCF) *Results-Oriented Accountability Plan* (ROA) identifies federal and state child welfare outcomes and the measures used to assess success in achieving these outcomes. Each measure is designed to collect a specific type of information, and these data are used to make conclusions about the attainment of target outcomes.

Data (plural for *datum*) are a collection of individual pieces of information about a variable of interest. A ***variable*** is a construct that varies; that is, it can take on more than one value. ***Categorical variables*** such as Race/Ethnicity and Gender do not have any inherent numerical properties. Values for these variables are often names of different groups; and although different numbers may represent these groups, these numbers do not indicate any quantitative differences between the groups. ***Continuous variables*** such as Age and Percentages do have numerical properties. Values for these variables are quantitatively relevant. For example, a two-year old child is twice as old as a one-year old child, and 80% is twice as much as 40%.

Statistics are a form of applied math that can be used to explain variables based on the data collected. Statistics can be divided into two types: ***descriptive statistics*** and ***inferential statistics***. Descriptive statistics *describe* a single variable based on the data that have been collected. Inferential statistics are used for testing hypotheses about relationships among multiple variables.

Descriptive statistics describe the characteristics of a sample or population. The most commonly used descriptive statistics are ***measures of central tendency*** and ***measures of variability***. Measures of central tendency describe the data in the middle (or central) part of the distribution of a continuous variable. Measures of variability help describe any differences between values that appear in the distribution of data for a specific variable. Descriptive statistics can often be displayed using tables, graphs, and charts.

Inferential statistics are powerful tools for understanding data and the relationships among different variables. Inferential statistics are based on hypothesis testing. Hypotheses are statements about the possible answers to the research question. There are always two hypotheses: generally speaking, one is a statement that the answer to the research question is *no* (this is called the *null hypothesis*), and the other is a statement that the answer to the research question is *yes* (this is the *alternative hypotheses*). Statistical tests allow us to infer if the answer to the question is *no*. There are three basic types of inferential statistics, and each measures a specific type of question: ***1) tests of association, 2) tests of group differences***, and ***3) tests of prediction.***

Introduction

On February 1, 2015, the Florida Department of Children and Families Office of Child Welfare released the *Results-Oriented Accountability (ROA) Plan*.[1] The ROA plan provides a "framework for measuring the success of efforts to improve Child Welfare outcomes while creating a culture of transparency and accountability" (p. 3). The *Cycle of Accountability* (Figure 1, ROA p. 8) consists of five phases, all of which rely in some way on the use of data and data analysis.

In order to monitor outcomes (Phase 1), data must be collected and analyzed (Phase 2). The results of the analyses are subjected to research review (Phase 3) for assessing external validity. Evaluation (Phase 4) assesses the internal validity of the results and implications for interventions. Quality improvement (Phase 5) is the implementation of new interventions.

This technical report provides an overview of key concepts involved in data analysis and includes demonstrations of the application of statistical techniques for data analysis. All data and statistical analyses in this report are based on the *Community-Based Care (CBC) Lead Agency Scorecard* for Quarter 3 (Q3) of Fiscal Year 2014-2015 (January 1, 2015 to March 30, 2015).[2]

Overview of Data Analysis Concepts

Making valid conclusions about child welfare outcomes is contingent on understanding the relationship between data and variables and how different statistical methods are needed based on the research questions.

What are Data?

Data (plural for *datum*) are a collection of individual pieces of information about a variable of interest. A *variable* is a construct that varies; that is, it can take on more than one value. For example, *Safety Measure #1* of the CBC (*Rates of abuse per 100,000 days in foster care*) is a variable because the rate can be different for different agencies. If all of the values for a variable are the same, then there is no variation in the values and the variable becomes a ***constant.***

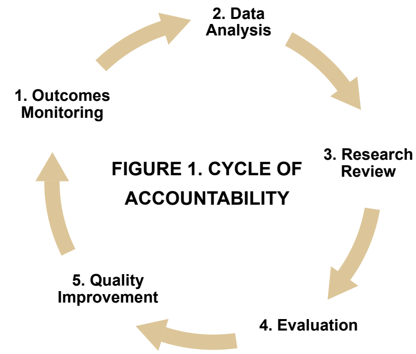

FIGURE 1. CYCLE OF ACCOUNTABILITY

Variables can be measured at different levels as well. ***Categorical variables*** such as Race/Ethnicity and Gender do not have any inherent numerical properties. Values for these variables are often names of different groups; and although different numbers may represent these groups, these numbers do not indicate any quantitative differences between the groups. For example, Gender might be coded as "1" = "male" and "2" = "female" in a data set. ***Continuous variables*** such as Age and Percentages do have numerical properties. Values for these variables are quantitatively relevant. For example, a two-year old child is twice as old as a one-year old child, and 80% is twice as much as 40%.

A series of measures have been developed as indicators of the child welfare outcomes identified in the ROA plan. Each measure is designed to collect a specific type of information, and these data are used to make conclusions about the attainment of target outcomes. The 12 measures on the CBC Scorecard are listed in Table 2. Additional information about the measures is available at http://www.dcf.state.fl.us/performance/cbc/CBC_Scorecard_Methodology.pdf

The data collected on these measures is used to answer questions about target outcomes. For example, one of the safety outcome measures is *percent of children who are not neglected or abused during in-home services*, and the standard is 95% will not be abused. Monitoring the safety outcome involves answering the question of whether or not the standard was met, such as whether or not each child in an agency was neglected or abused. These data will determine if an agency met the standard. Furthermore, data from multiple agencies can be aggregated to determine if the state as a whole met the standard.

Causality

One of the primary goals of research is to establish ***causality*** or *cause*–and–*effect* by identifying which variable(s) are responsible for changes in a target outcome. The *cause* is referred to as the ***independent variable****;* it is the variable that makes the outcome change. The outcome is referred to as the ***dependent variable*** because it is dependent on the independent variable. Causality is a key component of intervention research with the goal of demonstrating the success of the intervention, and the intervention alone, in achieving outcomes. For example, one component of the Attachment and Biobehavioral Catch-up (ABC) intervention is to help caregivers re-interpret children's behavior so that they can provide nurturance even when it isn't evoked.[3] One goal of the intervention is to "increase caregiver nurturance, sensitivity, and delight". Part of the evaluation of this intervention is whether or not it can be established that the intervention actually caused any positive outcomes for this goal (versus some other explanation of observed changes). Causality is the gold standard of intervention research but it is extremely difficult to establish outside of

Extreme caution should be used in making any claims of causality, and claims of causality should be carefully evaluated based on the required criteria.

controlled environments. Strict criteria must be met in order to establish causality. The first criterion is ***association***, meaning that there should be changes in the outcome measure when the intervention is used and no changes if the intervention is not used. The second criterion is ***solation***, meaning that any changes in the outcome measure can be attributed to the intervention and not another source. In order to meet this criterion, all other potential causes must be accounted for and controlled, a nearly impossible task in social science research. The third criterion is ***temporal precedence,*** which means the intervention has to be delivered before any changes in the outcome measure are observed. The fourth criterion is that there must be a ***theoretically plausible explanation*** for why the intervention would cause the target outcome to change. Extreme caution should be used in making any claims of causality, and claims of causality should be carefully evaluated based on the required criteria.

What are Statistics?

The word "statistics" can refer to many things, but statistics are always numbers that help us understand the data that have been collected. Statistics can be divided into two types: ***descriptive statistics*** and ***inferential statistics****.* Descriptive statistics *describe* a single variable based on the data that have been collected. Inferential statistics are used for testing hypotheses about relationships among multiple variables.

Part of a statistical analysis plan is determining what the ***unit of analysis*** will be. The unit of analysis is who or what the data represent. Although there is flexibility in choosing the unit of analysis, it is critical to specify the unit prior to data collection in order to ensure that the appropriate data are collected for that unit. In child welfare research it is common to see hierarchical, or multilevel, units of analysis. This occurs when data for one unit can be aggregated to become data for a different unit. For the CBC Scorecard, the data are presented for lead agencies, but agencies did not start out as the unit of analysis. First, data were collected on individual children, which means that *children* were the *unit of analysis* at that point and each piece of data belonged to an individual child. Data about individual children were then aggregated to create data about an agency. (This is the data shown on the CBC Scorecard). This process continues as data about agencies are then aggregated to create data about a state. The hierarchical nature of the data is demonstrated here as multiple individual children make up an agency and multiple agencies make up a state.

Populations and Samples

It is important to understand where data are collected from and what they represent. For example, were data collected on all families in the state of Florida or were data only collected from families in Orange County? Was the average number of families on a caseload collected from all agencies or only one agency? These decisions have implications for how data are used. In order to understand these implications, the "unit" of interest must be defined. As indicated before, this report is based on the definition of the "unit" of interests are people, for example investigators or supervisors. This collection of units can be defined as a "population" or a "sample".

Populations

A "population" consists of the entire collection of people (units) of interest. A population must be defined in order to understand who is part of the population and who is not. Defining a population means establishing criteria that clearly identifies who is in the population. Without the criteria it is impossible to know who is in the population. For example, suppose the population of interest is families receiving services in the state of Florida. This definition of the population alone is not sufficient; it does not say which families are part of the population. Clearly "receiving services" is a criteria, but does this mean any service or only certain services? What is the time frame? Is it only families currently receiving services, or is it families who received services between 2012 and 2014? There is not necessarily a right or wrong way to define a population, but it must be done in a way that matches what the data will be used for.

Samples

In order to ensure that data about a population are 100% accurate, data must be collected from *every* person (unit) in the population. There are many reasons why this may not be possible. For example, there are limited resources for collecting data, missing data about some individuals in the population, or inability to collect data from everyone in the population. In these instances the individuals who provided data are referred to as a "sample". A sample consists of a smaller group (or subset) of individuals that are a part of the entire population.

Descriptive Statistics

Descriptive statistics describe the characteristics of a sample or population, and in order to understand descriptive statistics, one must first understand how the data are distributed. A variable is measured by collecting data about it. Table 2 lists 12 measures from the CBC Scorecard and for each measure there are 20 pieces of data (corresponding to the 20 agencies). Descriptive statistics are used to explain the distribution of the data. Although there are several types of descriptive statistics, the two most commonly used and informative are ***measures of central tendency*** and ***measures of variability***.

Measures of Central Tendency

Measures of central tendency describe the data in the middle (or central) part of the distribution of a continuous variable. There are three common measures of central tendency: *Mean*, *Median*, and *Mode.*

The ***mean*** is the *average* of all of the values for a specific variable. It is created by adding up all of the values in the distribution and then dividing by the number of values.

The ***median*** (also called the *50th percentile*) is the value that is in the middle of the distribution, so that 50% of the values in the distribution are lower and 50% of the values are higher.

The ***mode*** is the value that appears the most often in the distribution.

Measures of Variability

Measures of variability describe differences in data collected for a variable. By definition a ***variable*** is something that varies or can take on more than one value. If all of the values in a distribution are the same, then there is no variability; instead, this would be called a constant. Measures of variability help describe any differences between values that appear in the distribution. There are many measures of variability and three common ones are listed here: *Frequency*, *Range*, and *Standard Deviation*.

The ***frequency*** reports how often each value for a variable occurs. Frequency distributions can be used for both categorical and continuous data.

The ***range*** refers to the difference between the lowest and highest values in the data.

The ***standard deviation*** indicates how close the values in the distribution are to the mean of the distribution. A small standard deviation indicates that the values in the distribution are very close to the mean value, whereas a large standard deviation indicates that values are more spread out from the mean value. The standard deviation can also help identify the percentage of scores that fall within a particular range. Based on the assumption of a normal distribution of values, approximately 68% of values fall within 1 standard deviation above and below the mean. Roughly 95% of values fall within 2 standard deviations above and below the mean. More than 99% of values fall within 3 standard deviations above and below the mean.

A small standard deviation indicates that the values in the distribution are very close to the mean value, whereas a large standard deviation indicates that values are more spread out from the mean value.

Applying Descriptive Statistics

Table 1 provides a summary of different types of descriptive statistics. Table 2 provides an array of descriptive statistics for 12 measures on the Community-Based Care Scorecard for FY 2014-2015 Q3.

TABLE 1. SUMMARY OF DESCRIPTIVE STATISTICS

Statistic	Description	Type of Data
Mean	Average of all values	Continuous
Median	50th percentile	Continuous
Mode	Most frequently occurring value	Continuous and Categorical
Frequency	How many times each value in the data set occurs	Continuous and Categorical
Range	Difference between the lowest and highest values in the data set	Continuous
Standard Deviation	Estimate of how far scores are from the mean	Continuous

TABLE 2. DESCRIPTIVE STATISTICS FOR 12 COMMUNITY-BASED CARE LEAD AGENCY SCORECARD MEASURES FOR FY 2014-2015 Q3

Measure	Domain	Mean	Median	Range	Standard Deviation
1. Rate of abuse per 100,000 days in foster care	Safety	9.58	9.31	5.1 – 14.9	2.95
2. % of children who are not neglected or abused during in-home services	Safety	96.9%	96.6%	92.7% - 100%	1.6%
3. % of children who are not neglected or abused after receiving services	Safety	96.7%	96.7%	92.1% - 100%	2.1%
4. % of children under supervision who are seen every 30 days	Safety	99.8%	99.9%	99.4% - 100%	0.2%
5. % of children exiting foster care to a permanent home within 12 months of entering care	Permanency	44.3%	45.7%	20.8% - 59.2%	9.2%
6. % of children achieving permanency in 12 months for children in foster care 12 to 23 months	Permanency	57.1%	57.7%	41.2% -72.0%	8.7%
7. % of children who do not re-enter foster care within 12 months of moving to a permanent home	Permanency	86.8%	87.3%	76.3% - 96.4%	5.6%
8. Children's placement moves per 1,000 days in foster care	Permanency	3.66	3.94	2.3 – 5.3	0.8%
9. % of children in out-of- home care who have received medical services in the last 12 months	Well-Being	96.9%	97.7%	89.5% - 100%	2.5%
10. Percent of children in out-of-home care who have received dental services in the last 7 months	Well-Being	90.4%	91.0%	80.0% - 97.2%	5.4%
11. % of young adults in foster care at age 18 who have completed or are enrolled in secondary education, vocational training, and/or adult education	Well-Being	86.4%	89.9%	58.0% - 100%	11.1%
12. % sibling groups where all siblings are placed together	Permanency	65.1%	76.1%	55.9% - 71.7%	5.3%

The mean was calculated for each measure by adding up each agency's data and then taking the average. So for item #2: on average, 96.9% of children were not neglected or abused during in-home services. Additional descriptive statistics show that there is some variability across agencies with the lowest value being 92.7% of children at an agency, and the highest value being 100% of children at an agency. Based on the standard deviation of 1.6, approximately 64% of the agencies should have results between 95.3% and 98.5% (calculated as the mean -/+ a standard deviation).

Descriptive statistics can also be provided visually through the use of charts and graphs. For example, data from the 20 agencies on the *Children's placement moves per 1,000 days in foster care* measure were plotted in a frequency distribution. When the data are continuous, as these data are for this measure, the graph is called a *histogram* (Figure 2). The distribution begins with the lowest value (2) on the left and then lists the data points in ascending order from left to right across the horizontal axis, ending with the highest value (5.5) on the right. This is the distribution of the data. What makes this a frequency distribution is that for each data point, the number of times that value appears in the data is marked by the vertical axis. For example, there are seven agencies reporting between 4.0 and 4.5 moves and only one agency reporting between 3.5 and 4.0 moves.

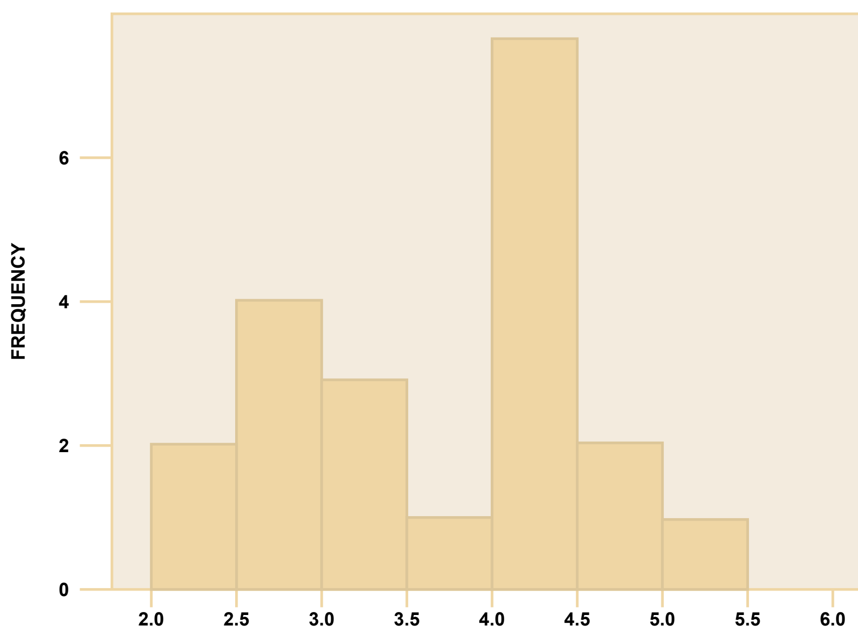

FIGURE 2. FREQUENCY DISTRIBUTION FOR CHILDREN'S PLACEMENT MOVES PER 1,000 DAYS IN FOSTER CARE

Similarly, frequency distributions can be created for categorical data. These are called bar charts. The 20 agencies on the Scorecard were grouped by region (Southern, Central, or Northern), and then the means for three measures were calculated for each region. Figure 3 displays the mean values of the three measures for each region separately.

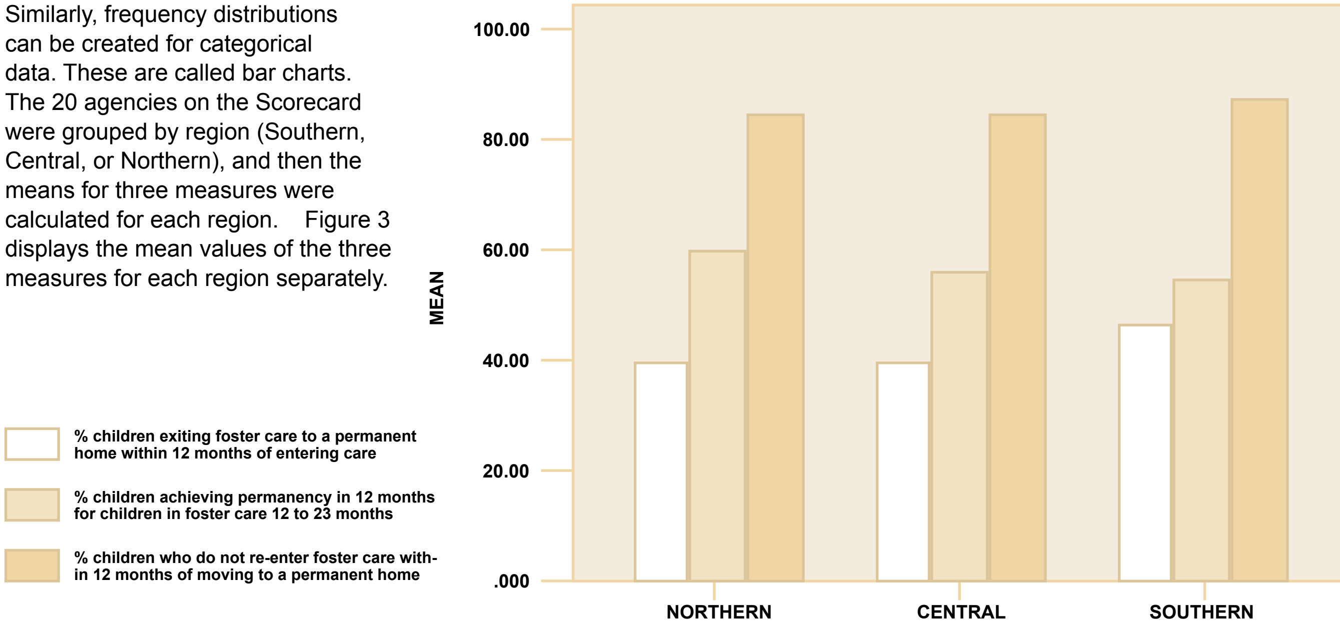

FIGURE 3. PERMANENCY OUTCOMES BY GEOGRAPHIC REGION

Another way to display this data is a *pie chart*. A pie chart can be used to show the number and/or percentage of observations that fall into a set of categories. Figure 4 displays the percentage of agencies that were below target, at target, or above target for item #10, related to receiving dental services. Approximately 40% of agencies (*n*=8) were below the target of seven months, whereas only 25% of agencies (*n*=5) were above the target or seven months.

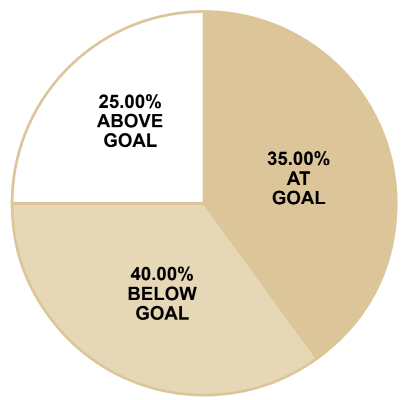

FIGURE 4. GOAL ATTAINMENT FOR % OF CHILDREN IN OUT-OF-HOME CARE WHO HAVE RECEIVED DENTAL SERVICES IN PAST 7 MONTHS

Inferential Statistics

Inferential statistics are powerful tools for understanding data and the relationships among different variables. This type of statistic is used to test hypotheses and make decisions about whether or not results from analyzing data from a sample of people are likely to be the same for the whole population of people. Understanding the use of inferential statistics requires understanding hypothesis testing.

As noted previously in this report, only research questions about the relationships among multiple variables require statistical tests. For example, answering the question, "*What is the percent of children who are not neglected or abused during in-home services" (Safety Measure #1*) does not require a statistical test. This is a single variable and the answer can be derived using descriptive statistics.

In contrast, questions about multiple variables do require statistical testing. For example, is the *% of children under supervision who are seen every 30 days (Time Standard #4)* related to the *% of sibling groups where all siblings are placed together (Well-Being Measure #12*)?

This type of research question has hypotheses. Hypotheses are statements about the possible answers to the research question. There are always two hypotheses: generally speaking, one is a statement that the answer to the research question is *no* (this is called the *null hypothesis*), and the other is a statement that the answer to the research question is *yes* (this is the *alternative hypotheses*). Statistical tests allow us to infer if the answer to the question is *no*.

It is straightforward to answer a research question when there is data on everyone in the population of interest, and this is potentially the case when using statewide data. However, sometimes it isn't possible to have data on every case and a sample has to be used instead. When a sample is used, there is a possibility that the people in the sample are in some way different from the people in the whole population, and this is called ***sampling error***. For example, is there the potential for sampling error if data from Pasco, Pinellas, and Hillsborough Counties were used to make a conclusion about statewide outcomes? Or is there the potential for sampling error if estimating annual statewide outcomes using only data from one quarter? When this happens, any observed results may be due to chance instead of being real. It would be a mistake to say that the answer to the research isn't *no*, when in fact the answer really is *no*. Statisticians refer to this as a *Type 1 error*. Similarly, it would be a mistake to say that the answer to the research question is *no* when it actually isn't *no*. Statisticians refer to this as a *Type 2 error*. In general, statistical analyses are designed to avoid making Type 1 errors.

There is always a chance of making a mistake when applying the results from a sample to the entire population. Therefore data analysts want to be confident they are not committing a Type 1 error. The most commonly used criterion is a confidence level of 95%, meaning there is a 95% chance of *not* committing a Type 1 error. Using this level of confidence means that there can be no more than a 5% chance of drawing the wrong conclusion. These two values can vary depending on the desired level of confidence, but they must always add up to 100%. For example, if you wanted to be 99% confident, then there can be no more than a 1% chance of committing a Type 1 error. The probability of committing a Type 1 error is referred to as the *alpha*-level (the symbol is α), and the researcher chooses it.

Statistical tests produce ***p-values***. The *p*-value is the actual probability of committing a Type 1 error based on the actual data. Deciding if the answer to the research question is *no* or *not no* depends on the relationship between the alpha-level and the *p*-value. As long as the actual chance of committing a Type 1 error (the *p*-value) is less than the maximum allowable chance (the alpha-level), then it is possible to infer that the answer to the research question is *no.* This is referred to as ***statistical significance***.

There are three classes of inferential statistics, and each measures a specific type of question. A summary of inferential statistics is provided in Table 3 on page 9.

Tests of Association

These types of tests answer the question of whether or not two or more variables are associated in some way. ***Associated*** in this context means that the values for one variable are somehow connected to the values of another variable. As the values for one variable change, the values for another variable are also changing. If the variables are continuous, it is possible that the variables increase or decrease at the same time; this is referred to as a ***positive correlation***. It is considered *positive* because the variables are changing in the same direction. Conversely, ***negative correlations*** refer to situations where the variables are changing in opposite directions. Technically, tests of associations do not have independent and dependent variables, although sometimes it is possible to identify which variable might be influencing the other variable.

The term ***correlation*** refers to the relationship between the variables and indicates how strong the association is and which direction it goes positive or negative. Correlations can range from -1 to +1, and the further the magnitude is from 0, the stronger the association. If the variables are categorical, then a different statistic called chi-squared is used. (*Chi* is a Greek letter and is pronounced like “Hi” but with a “K”). Tests of associations allow one to make inferences about whether or not the variables are connected in some way. Just because there is an association between variables does not mean that one variable is *causing* the other variable. As noted in the beginning of this report, there are strict criteria for claiming that one variable causes another variable. ***Association*** is a required element of causality but is not sufficient on its own to establish causality.

Association is a required element of causality but is not sufficient on its own to establish causality.

Applying Tests of Association

Question #1: Is there a positive association between the *% of children who are not neglected after receiving services (Safety Measure #3)* and the *% of children exiting foster care to a permanent home within 12 months (Permanency Measure #5)*?

Answer #1: Although the correlation (*r*) between the variables is positive as hypothesized (.20), the probability of committing a Type 1 error is high (p = .39). Therefore, based on the data available for the specified time period, the answer to Question 1 is *no.*

Question #2: Is there a negative correlation between the *average placement moves per 1,000 days (Permanency Measure #7)* and the *% of children achieving permanent placement within 12 months in foster care 12-23 months (Permanency Measure #8)*?

Answer #2: There is a strong negative correlation (r = -.71) between the two variables. The *p*-value for this analysis is <.001, indicating that there is less than .1% probability of committing a Type 1 error. Given that the acceptable level is 5%, based on the data available for the specified time period, the answer to Question 2 is *yes*.

Tests of Group Differences

These types of tests answer the question of whether or not two or more groups (the independent variable) are different with respect to a dependent variable. Tests of group differences always focus on the mean value of the variable of interest for each group and statistical tests help determine if any observed differences between groups are real or only due to chance. In this analysis, the result is reported as the magnitude of the difference(s) and which groups are higher/lower. ***Analysis of Variance (ANOVA)*** is the statistical test of group differences when there are 3 or more groups involved. A *t-test* is used if there are only 2 groups. The basis of tests of group differences is to determine if the differences between groups is greater than any differences within the groups themselves. That is, the analysis determines if the difference (*variance*) between groups is real or simply due to chance because of sampling error or other factors. In addition to identifying differences between groups, these tests will also determine which groups are actually different.

Applying Tests of Group Differences

Question #1: Are there differences in the *% of young adults in foster care at age 18 who have completed or are enrolled in secondary education, vocational training, and/or adult education (Well-Being Measure #11)* based on geographic region of the agencies (Northern, Central, or Southern)?

Answer #1: Based on the results of the ANOVA, there is a statistically significant difference between at least two of the regions. The *p*-value for this analysis is .03, indicating that there is a 3% probability of committing a Type 1 error. Given that the acceptable level is 5%, based on the data available for the specified time period, the answer to Question 1 is *yes*. However, when there are more than three groups being compared, additional statistical test have to be performed in order to know which groups specifically are different. These secondary tests are referred to as post hoc analyses. The results of the post hoc analyses reveal that there is a statistically significant difference between the central and southern regions, but not between the northern region and either of the other two. Based on this data, the southern region had 13.94% of young adults achieving this goal than the central region.

Question #2: Are there differences in the *% of children who are not neglected or abused after receiving services (Safety Measure #3)* between agencies with *high versus low rates of abuse per 100,000 days in foster care (Safety Measure #1)*?

Answer #2: Based on the results of the t-test there is a significant difference in *% of children who are not neglected or abused after receiving services (Safety Measure #3)* based on high versus low *rates of abuse per 100,000 days in foster care (Safety Measure #1)*. The *p*-value for this analysis is .02, indicating that there is a 2% probability of committing a Type 1 error. Given that the acceptable level is 5%, based on the data available for the specified time period, the answer to Question 2 is *yes*. On average, agencies with low rates scored 2.02% higher.

Tests of Prediction

These types of tests answer the question of whether or not a group of independent variables can predict the value of a dependent variable. Many of these tests fall under the umbrella term of ***regression analyses***. Regression analyses are powerful tools that allow analysts to anticipate outcomes. Another benefit of regression analyses is the inclusion of multiple predictors that allow the analyst to estimate the relationship between a predictor and an outcome after controlling for the influence of other predictors. ***Linear regression*** is the statistical analysis used to predict values for continuous variables. The results of a linear regression analysis indicate the magnitude and direction of predicted change in the outcome each time the predicting variable changes by 1 point. These estimated changes can be expressed in the original metric of the predictor (e.g., predicted change in % of the outcome) for an increase of 1% of the predictor, called an ***unstandardized coefficient*** or as standard deviations (e.g., predicted change in standard deviations of the outcom)e for a 1 standard deviation change in the predictor, called a ***standardized coefficient***. Both types of results are informative, and the decision to use one or both of them is often based on the way the outcome variable is measured. For example, if the outcome is measured in %, it can only be between 0% and 100%. It is possible that the unstandardized coefficient predicts a (non-existent) value greater than 100%; in this case using standardized coefficients is better.

Tests of prediction can also be applied to outcomes that are categorical. ***Binomial regression*** is the statistical analysis used to predict a categorical outcome when there are only two possible outcomes; for example, a goal is met or not met. ***Multinomial regression*** is used when there are more than two possible outcomes for a categorical variable; for example, goal met, goal partially met, or goal not met.

Applying Tests of Predictions

Question #1: Can the variables *% of children under supervision seen every 30 days* and *% of children exiting foster care to a permanent housing within 12 months* predict well-being outcomes?

Answer #1a: Based on the results of the linear regression analysis, both of these variables are statistically significant predictors of *% of children in out-of-home care who receive medical service in the last 12 months*. Using standardized coefficients, the *% of children in out-of-home care who receive medical service in the last 12 months* is predicted to increase by .53 standard deviations for each 1 standard deviation increase in *% of children under supervision seen every 30 days* controlling for the other predictor. The *% of children in out-of-home care who receive medical service in the last 12 months* is predicted to increase by .43 standard deviations for each 1 standard deviation increase in *% of children exiting foster care to a permanent housing within 12 months* controlling for the other predictor. The p-value for both predictors is <.02, indicating that there is less than a 2% probability of committing a Type 1 error. Given that the acceptable level is 5%, the conclusion is based on the data available for the specified time period; the answer to Question 1 is *yes*.

Answer #1b: Based on the results of the linear regression analysis, one of these variables is a statistically significant predictor of % of children in out-of-home care who have received dental services in the last 7 months, and one of the variables is not. Using standardized coefficients, the % of children in out-of-home care who have received dental services in the last 7 months is predicted to increase by .45 standard deviations for each 1 standard deviation increase in % of children under supervision seen every 30 days controlling for the other predictor. The p-value for this predictor is .046, indicating that there is a 4.6% probability of committing a Type 1 error. Given that the acceptable level is 5%, based on the data available for the specified time period, the answer to Question 1 is *yes* for this predictor. The % of children in out-of-home care who have received dental services in the last 7 months is predicted to increase by .28 standard deviations for each 1 standard deviation increase in % of children exiting foster care to a permanent housing within 12 months controlling for the other predictor. However, the p-value for this predictor is .20, indicating that there is a 20% probability of committing a Type 1 error. Given that the acceptable level is 5%, based on the data available for the specified time period, the answer to Question 1 is *no* for this predictor.

TABLE 3. SUMMARY OF SELECT INFERENTIAL STATISTICS

Statistical Analysis	Description	Independent Variable	Dependent Variable
Correlation	Association between 2 continuous variables	NA	NA
Chi-square test of independence	Association between 2 categorical variables	NA	NA
t-Test	Difference in outcome between two groups	Groups (Categorical)	Outcome (Continuous)
ANOVA	Difference in outcome between 3+ groups	Groups (Categorical)	Outcome (Continuous)
Linear Regression	Predicted change in outcome	Predictors (Continuous/Categorical)	Outcome (Continuous)
Binomial Regression	Odds of an outcome	Predictors (Continuous/Categorical)	Outcome (Categorical; Dichotomous)
Multinomial Regression	Odds of an outcome	Predictors (Continuous/Categorical)	Outcome (Categorical; Multinomial)

Summary

Data and data analysis are critical components of ongoing outcome assessment for child welfare policy and practice. As demonstrated in the cycle of accountability, data are an integral part of each phase of outcome assessment. This process requires evidence in the form of data collected though identified measures. High quality measures result in high quality data, and as data quality increases, so does the confidence in the validity of the conclusion based on the data. Although not addressed in this report, measurement quality is covered in depth in the *Results-Oriented Accountability (ROA) Plan.*

High quality measures result in high quality data, and as data quality increases, so does the confidence in the validity of the conclusion based on the data.

Analysts have an array of statistical tools at their disposal for making sense of large amounts of data collected through the child welfare data reporting system, but data can be used at many levels of the child welfare system. As shown on the Community-Based Care Lead Agency Scorecard, individual agencies can use data to describe their own progress in meeting target outcomes. This hierarchy of data extends both ways. More specific data levels within agencies might include sub-agencies of the lead agency, departments within an agency, or even caseloads of employees within a department. More aggregated data levels might include counties, regions, or state(s).

Regardless of the data level, descriptive and inferential statistics can be used to make sense of the data. Statistical results can be used to understand what has happened in the past, what is happening in real time, and what might happen in the future. Data collection can be a high-resource activity requiring substantial time and attention to be conducted correctly. However, collecting and analyzing appropriate data will lead to improved child welfare outcomes and improve the safety, permanency, and well-being of children in the state of Florida.

References

[1] http://www.dcf.state.fl.us/programs/childwelfare/docs/2015LMRs/Results-Oriented%20Accountability%20Plan.pdf

[2] http://www.dcf.state.fl.us/performance/cbc/CBC%20Scorecard%20SFY2014-2014%20Quarter%203%2020150416.pdf

[3] http://www.cebc4cw.org/program/attachment-and-biobehavioral-catch-up/