

## Appendix A

# Statistical Analysis

**This page is intentionally blank.**

## Statistical Analysis (Appendix A)

### 1.0 Introduction

Associated with each performance measure in the BellSouth OSS test is either a parity standard (for example, “ALEC orders are provisioned within the same time intervals as those for BellSouth retail”) or a benchmark standard (for example, “90% of ALEC address verification requests are returned within 10 seconds”).

When a parity standard applies, KPMG Consulting tested whether the performance measured in the test data is equal to the retail analog performance. When a benchmark standard applies, KPMG Consulting evaluated the test data against the fixed benchmark.

The statistical testing is subject to two types of statistical errors, Type I error and Type II error. A Type I error occurs if BellSouth fails a measure when it should have passed. A Type II error occurs if BellSouth passes a measure when it should have failed. Both benchmark and parity tests are subject to these types of error.

### 2.0 Statistical Methodology

The statistical methodology had several key components. First, Null and Alternative Hypotheses were established. Next, target Type I and Type II error rates were established. Finally, the evaluation method was established. The evaluation method specifies the exact statistical test to be performed.

#### 2.1 Null and Alternative Hypotheses

A standard statistical hypothesis-testing framework was used in the BellSouth test. The two mutually exclusive hypotheses in the BellSouth evaluation were:

- ◆ Null Hypothesis: BellSouth is meeting or exceeding the standard.
- ◆ Alternative Hypothesis: BellSouth is not meeting the standard.

#### 2.2 Test Error Levels

The Type I error was limited to 5% in this test.<sup>424</sup> This limit allows the Type II error to vary with sample size. In order to ensure that the Type II error was small, KPMG Consulting worked with the Florida Public Service Commission in advance of the test to ensure the precision for each Service Quality Measurement (SQM) is better than 20%.<sup>425</sup>

---

<sup>424</sup> This Type I error applies regardless of whether the standard is a parity standard or the standard is a benchmark standard. Statistical tests are not applied for the purposes of ongoing monitoring of benchmarks, but there is a distinction in purpose between the OSS test and ongoing monitoring efforts. Ongoing monitoring efforts may determine whether BellSouth is performing below a standard for a specific set of data. The OSS test seeks to determine whether the test outcomes were consistent with an OSS that is generally operating at or above an acceptable level. As such, random variation in test outcomes is necessarily considered in the OSS test, via statistical testing, regardless of whether the appropriate standards are benchmarks or parity measures.

<sup>425</sup> This analysis was performed using BellSouth data and applies to sample sizes in the major OSS testing areas. Precision was defined as the ratio of the standard error for the measure to the average for that same measure. In tests that KPMG Consulting has performed in some other jurisdictions, a minimum sample size of 140 was established for certain measures. This sample size ensures that Type II error is 5% if the difference for a parity measure is .28 standard deviations. It also ensures that the Type II error rate is less than 5% for a benchmark standard of 90%, if the true performance is at 80%. While these considerations were not part of the specific design for the Florida test, the sample

### 3.0 Statistical Evaluation Methods

The following tables summarize the evaluation method used for each type of measure. Note that for parity tests, two samples were compared, that of the KPMG Consulting test data and that of the BellSouth retail data.

**Table A-1: Parity Test Evaluation Methods**

Type Of Test	Rate or Interval	Sample Sizes
Permutation Test	Interval	Either sample < 200
Modified Z-Test	Interval	Both samples $\geq$ 200
Hypergeometric	Rate	Both samples < 10,000
Binomial	Rate	Either sample $\geq$ 10,000
Poisson	Rate – not proportion	NA

**Table A-2: Benchmark Test Evaluation Methods**

Type Of Test	Rate or Interval	KPMG Consulting Test Sample Size
Median Test (using Binomial)	Interval	<200
One sample t-test	Interval	$\geq$ 200
Binomial	Rate	Any
Poisson	Rate – not proportion	Any

### 4.0 Description of Specific Evaluation Methods

Each of the tests listed above, with the exception of the modified z-test, is a standard statistical test. A description of these tests follows.

For parity tests of intervals, KPMG Consulting used a modified z-test for services/products where the sample size is greater than, or equal to, 200 for both the BellSouth retail and KPMG Consulting test data. For small samples (when one sample is less than 200), a permutation test was used. A permutation test does not make implicit assumptions about the probability distribution of the underlying data.

A modified z-test is similar to a two-sample t-test. Like the pooled variance version of the two-sample t-test, the modified z-test assumes, under the Null Hypothesis, that the BellSouth retail and the KPMG Consulting test data have equal variances. A modified z-test also assumes a large enough sample size to allow distributional assumptions of the test to be ignored. In particular, the modified z-test assumes the data come from a particular probability distribution called the Normal distribution. This assumption is practical for large sample sizes, because the distribution of the average of a large sample is close to a Normal distribution. The modified z-test only uses the BellSouth retail sample variance, not the pooled variance. The result is a test with greater power for testing against alternatives where the KPMG Consulting test data variance is higher.

---

sizes in the Florida test meet the minimum sample size criteria used in these other tests.

For parity tests of rates, a Hypergeometric test was used when sample sizes are less than 10,000 for both the BellSouth retail and KPMG Consulting test data. The Hypergeometric test allows for an exact measurement of the statistical probabilities for Type I and Type II errors. When either the BellSouth retail or KPMG Consulting test sample size is greater than or equal 10,000, a Binomial test was used. The Binomial test assumes the BellSouth retail proportion is exact, but will not affect the test results for large samples. While using a Binomial test instead of a Hypergeometric test could result in a different outcome, KPMG Consulting found no cases in the test data where such a difference in outcome existed.

For benchmark tests for intervals, a one-sample t-test was used for sample sizes above 200. For sample sizes below 200, a Binomial test was used, and the Null Hypothesis assumed the median of the data equals the benchmark.

For benchmark tests for rates, a Binomial test was used. The Binomial test allows for an exact measurement of the statistical probabilities for Type I and Type II errors.

**This page is intentionally left blank.**