BEFORE THE

FLORIDA PUBLIC SERVICE COMMISSION

DIRECT TESTIMONY OF

ROBERT M. BELL, PH.D.

ON BEHALF OF

AT&T COMMUNICATIONS OF THE SOUTHERN STATES, INC.
WORLDCOM, INC.
DIECA COMMUNICATIONS COMPANY D/B/A COVAD COMMUNICATIONS COMPANY
NEW SOUTH COMMUNICATIONS CORP.
MPOWER COMMUNICATIONS CORP.
E.SPIRE COMMUNICATIONS, INC.
ITC^DELTACOM COMMUNICATIONS, INC.
RHYTHMS LINKS INC.

DOCKET NO. 000121-TP

MARCH 1, 2000

Q. PLEASE STATE YOUR NAME AND BUSINESS ADDRESS.

A. My name is Robert M. Bell. My business address is AT&T Labs-Research, 180 Park Avenue, Florham Park, New Jersey 07932.

Q. PLEASE DESCRIBE YOUR EDUCATIONAL AND PROFESSIONAL BACKGROUND.

A. I received a Ph.D. in Statistics from Stanford University in 1980. From 1980 to 1998, I worked as a statistician at RAND, a non-profit institution that conducts public-policy analysis. While at RAND, I supervised the design and/or analysis of large multi-site evaluations in the fields of preventive dentistry, drug prevention, and depression care. I also headed the RAND Statistics Group from 1993 to 1995 and taught statistics in the RAND Graduate School from 1992 to 1998. Since 1998, I have worked in the Statistics Research Department at AT&T Labs-Research. I have authored or co-authored 50 refereed articles on statistical analysis that have appeared in a variety of professional journals. I am a fellow of the American Statistical Association. I currently serve on the Panel to Review the 2000 Census organized by the National Academy of Sciences.

Q. WHAT IS THE PURPOSE OF YOUR TESTIMONY?

A. My testimony describes the statistical methodology the Florida Public Service Commission should adopt for use in comparing BellSouth's performance provided to itself and its affiliates with the performance it provides to Alternative Local Exchange Companies (ALECs). I also recommend an appropriate range for the parameter delta used in connection with the statistical methodology. (Issues 11.C.1 and 2; 12.C.1 and 2)

Q. WHY ARE STATISTICAL TESTS USEFUL TOOLS?

A. Merely reporting averages of performance measurements alone, without further analysis, does not indicate whether differences in performance results for ALEC customers versus a retail analog reflect actual discrimination or simply random variation. Once appropriate measures and comparison samples have been established, statistical tests compare the size of observed differences with the amount that could be expected to occur by chance under conditions of true parity of service. These comparisons help to determine quantitatively whether BellSouth has provided nondiscriminatory treatment to ALECs for measures with a retail analog. The FCC supported the use of statistical comparisons in its Bell Atlantic Order for New York. See *In the Matter of Application of Bell Atlantic for Provision of In-Region InterLATA Services In New York*, CC Docket No. 99-295 (December 23, 1999), Appendix B, Para. 2&4, where FCC stated:

> When making a parity comparison, statistical analysis is a useful tool to take into account random variations in

> the metrics. In the Second BellSouth Louisiana Order, we encouraged BOCs to submit data allowing us to determine if any detected difference between the wholesale and retail metrics is statistically significant.

Q. WHAT SHOULD THIS COMMISSION ORDER CONCERNING THE STATISTICAL METHODOLOGY TO BE USED IN EVALUATING PARITY?

A. There are two things that should be included in the Commission's order. First, the Commission should select the appropriate statistic for making parity determinations. My testimony explains that the modified z is the most appropriate statistic for this purpose. Second, the Commission should set the value of a parameter "delta," which is needed to compute a balancing critical value, at no higher than 0.25 for all submeasures. The modified z statistic compared with a balancing critical value based on a parameter delta no higher than 0.25 for all submeasures will enable this Commission to detect lack of parity in BellSouth's performance to ALECs.

Issue 11. a. What is the appropriate methodology that should be employed to determine if BellSouth is providing compliant performance to an individual ALEC? (Tier 1)

c. What is the appropriate structure?

1. What is the appropriate statistical methodology?

Issue 12. a. What is the appropriate methodology that should be employed to determine if BellSouth is providing compliant performance to an individual ALEC? (Tier 2)

c. What is the appropriate structure?

1. What is the appropriate statistical methodology?

Q. WHAT STATISTICAL METHODOLOGY DO THE ALECS RECOMMEND?

A. The ALECs recommend use of the modified z statistic. This test statistic is described in a paper attached to this testimony as Exhibit RMB-1.[1] For each parity submeasurement (a disaggregated measure), BellSouth's performance for its retail operation (or that of its affiliates) is compared with the performance it provides to a given ALEC to create a z score (the modified z statistic), which then can be used to determine whether BellSouth's performance for an ALEC is in parity with its performance for its retail operation. For small sample sizes (30 or fewer observations in either of the data sets to be compared), permutation analysis is used to compute the z score. Permutation analysis is a computer-intensive method that compares the observed results for the ALEC customers with the distribution of

[1] *See* Exhibit RMB-1, "Statistical Tests for Local Service Parity", Version 1.0, February 6, 1998, Local Competition Users Group.

results that would be observed if ALEC had been drawn at random from the pool of ALEC and BellSouth customers (see Exhibit RMB-2, "Permutation Analysis Procedural Steps").

Out-of-parity performance occurs when the z score falls below a pre-specified critical value that depends on the two sample sizes. Values of z that fall below the critical value are taken as indications of discrimination. The ALECs use a principle called "balancing" to determine the critical value.

Q IS MODIFIED Z AN APPROPRIATE COMPONENT OF THE STATISTICAL METHODOLOGY FOR MAKING PARITY DETERMINATIONS?

A. Yes. Experience with BellSouth's raw data confirms that the modified z statistic is an appropriate and effective component of the methodology for parity determinations. In its August 31, 1998 order in Docket No. U-22252-C, the Louisiana Public Service Commission required BellSouth to give ALECs access to raw data that underlies BellSouth's reports.[2] In that proceeding, Dr. Colin Mallows, an AT&T statistician, was able to receive and work with at least some of BellSouth's performance data in order to assess the performance of the statistical test.[3] The ability to look at and analyze data is critical to

[2] Order, In re: BellSouth Telecommunications, Inc., Service Quality Performance Measurements, Docket No. U-22252, Subdocket C, August 31, 1998.
[3] Pursuant to a protective agreement, BellSouth provided some of its raw data associated with four measures it includes in its SQM. The measures for which Dr. Mallows received

determining the appropriate statistical test; one cannot be assured that the data characteristics are properly accounted for in the statistical methodology unless one can observe the data and how it behaves over time. The Louisiana Public Service Commission's order provided the opportunity for Dr. Mallows to actually see raw data and, thereby, confirm and refine the statistical methodology. Dr. Mallows analysis of the raw data confirmed that the modified z statistic is an effective component of the methodology for parity determinations.

Q. WHAT IS THE CRITICAL VALUE AND WHY IS IT IMPORTANT?

A. The critical value is used, along with the modified z, to determine whether the performance for a particular measure is considered to be in violation. As the modified z statistic is defined in the ALEC plan, negative values of modified z provide evidence than an ALEC's customers are receiving worse service than the corresponding CLEC customers, with large negative numbers providing the most evidence. The value of the modified z statistic is compared with a pre-specified negative number, called the critical value. If modified z is more negative than the critical value, then the measure is determined to be in violation. Otherwise, the measure is not determined to be in violation, even though service for the ALEC customers may have been worse than service received by the retail customers.

---

some raw data were: Order Completion Interval, Maintenance Average Duration, Missed Repair Appointments, and Missed Installation Appointments.

Q. PLEASE EXPLAIN THE CONCEPT OF "ERROR" IN CONNECTION WITH STATISTICAL TESTING.

A. Because statistical tests are based on finite amounts of data, they are subject to error. For tests of parity, there is some chance that a measure will be determined in violation when, in fact, the two processes were in perfect parity (i.e., any difference was purely due to random variation). Likewise, when the two processes are out of parity such that the ALEC's customers receive systematically worse service, there is a chance that the statistical test will fail to find the measure in violation, again due to random variation.

Q. WHAT IS A TYPE I ERROR?

A. A Type I error occurs if the statistical test indicates that BellSouth is favoring its retail operations when, in fact, parity service exists. Type I errors occur because of random variation.

Q. WHAT IS A TYPE II ERROR?

A. A Type II error occurs if the statistical test fails to indicate that BellSouth is favoring its retail operations when, in fact, a certain degree of disparity does exist. Like Type I errors, Type II errors occur because of random variation. In contrast to Type I errors, determination of the probability of a Type II error requires specification of an alternative hypothesis that quantifies the degree of service disparity.

Q. HOW DOES THE CHOICE OF THE CRITICAL VALUE AFFECT TYPE I AND TYPE II ERRORS?

A. The critical value trades off between the probabilities of Type I and Type II errors. A large negative critical value holds down the probability of a Type I error, but allows the probability of a Type II error to grow larger. A less negative critical value keeps down the probability of a Type II error but allows the probability of a Type I error to grow. Put simply, a large negative critical value reduces the possibility of determining noncompliance when BellSouth is in fact providing parity service, while less negative values reduce the possibility of determining BellSouth is compliant when in fact they are providing noncompliant support.

Q. WHAT IS A BALANCING CRITICAL VALUE AND HOW IS IT DETERMINED?

A. The balancing critical value method explicitly accounts for the probabilities of both Type I and Type II errors. The basic concept is to equate the probability of a Type I error (under parity) with the probability of a Type II error for a specified alternative hypothesis.

Q. HOW SHOULD THE ALTERNATIVE HYPOTHESIS FOR THE BALANCING METHOD BE DETERMINED?

A. The alternative hypothesis should describe the minimum degree of disparity that constitutes a material impact on competition. The balancing method recognizes that small degrees of disparity may not significantly hinder competition, and thereby do not require protection for the ALECs. However, the degree of disparity specified by the alternative hypothesis should not exceed the minimum amount that would constitute a material impact on competition because doing so would deny the ALECs adequate protection against that degree of discrimination.

Q. HOW IS A BALANCING CRITICAL VALUE DETERMINED?

A. Once the alternative hypothesis has been specified, a balancing critical value (BCV) is set by equating the probabilities of Type I and Type II errors. The degree of disparity can be specified in terms of a parameter "delta."

Issue 11. a. What is the appropriate methodology that should be employed to determine if BellSouth is providing compliant performance to an individual ALEC? (Tier 1)

c. What is the appropriate structure?

2. What is the appropriate parameter delta, if any?

5. Should there be a floor on the balancing critical value?

Issue 12. a. What is the appropriate methodology that should be employed to determine if BellSouth is providing compliant performance to an individual ALEC? (Tier 2)

c. What is the appropriate structure?

2. What is the appropriate parameter delta, if any?

5. Should there be a floor on the balancing critical value?

Q. WHAT IS THE PARAMETER "DELTA" AND WHY IS IT IMPORTANT?

A. The parameter delta defines the degree of violation of parity (i.e., the alternative hypothesis) for which the probability of Type II error is balanced against the probability of Type I error under parity. Delta specifies the difference between the ALEC mean and the BellSouth mean. To account for the fact that performance measures do not share a common scale, the difference between the ALEC and BellSouth means is stated as delta times the standard deviation for BellSouth customers. For example, suppose that the measure Order Completion Interval has a mean of 5.0 days and a standard deviation of 6.0 days for BellSouth customers. Then a delta of 0.25 would yield an alternative hypothesis that the true mean for ALEC customers is 6.5 days (5.0 + 0.25 x 6.0).

Q. HAS A VALUE OF THE DELTA PARAMETER BEEN AGREED UPON?

A. No, the ALEC's and BellSouth's statisticians agree on the principle of balancing Type I and Type II errors, but they have not agreed on a value for the delta parameter. The balancing critical value development is incomplete until the value of the delta parameter is specified.

Q. WHY HAS THE DETERMINATION OF THE DELTA PARAMETER NOT BEEN RESOLVED?

A. Resolution of this question cannot be based solely on a theoretical statistical analysis. Ideally, this decision should be based on business judgment, namely by determining the smallest violation of parity that is "material." The parameter delta measures the size of this violation. Once delta is chosen, the formula makes proper allowance for the effect of the sample size. When delta is large, the balancing occurs at a more extreme degree of observed disparity. BellSouth wants a large delta because this means a smaller probability of Type I error and hence, larger probability of Type II errors for any given degree of true disparity. The ALECs want a value of delta that protects them against any degree of disparity that would pose a material obstacle to competition. If the parameter delta is set too high—such that some smaller violation would present a material obstacle to competition—then the balancing principle would be violated.

Q. WHAT VALUE OF DELTA DO THE ALECS PROPOSE AND WHAT DOES THAT VALUE IMPLY?

A. The ALECs propose that this Commission adopt 0.25 or less as the parameter delta value for all submeasures. To understand the implications of this and various alternative values of delta, consider what they imply for the counted performance measures. Consider a counted measure indicating a particular service problem that is triggered for 1 percent of BellSouth's own customers. Column 1 of Table 1 (below) shows that the degree of disparity quantified by delta equal to 0.25 implies that 5.0% of ALEC customers would encounter the same problem; that is, the ALEC rate is five times the BellSouth rate.[4] Subsequent rows of the same column show the problem rates for ALEC customers implied by a delta of 0.25 for problems that affect 5, 10, or 20 percent of BellSouth customers. The ALECs judge that disparities of this size pose material obstacles to competition. Therefore, delta should be no more than 0.25. Any larger value of delta would require even greater disparities before balancing takes place. For example, for a problem that occurs for 1 percent of BellSouth customers, a delta value of 0.50 would not balance until the ALEC rate reached 11.8%, nearly a twelve-fold increase. These disparities are highlighted in Table I.

Table 1

Percentage of ALEC Customers Receiving Bad Service, by BellSouth Percent and Delta

	Delta		
BellSouth Percent	0.25	0.50	1.00
1.0	5.0	11.8	31.9
5.0	11.8	21.0	44.0
10.0	18.7	29.3	53.6
20.0	30.8	42.8	67.4

Q. WHAT ARE THE CONSEQUENCES IF DELTA IS SET TOO LARGE?

[4] The table assumes use of arcsin square root transformation to stabilize the variance of proportions. Using this function, transformed proportions have a nearly constant variance across the range of possible true proportions.

A. Suppose that delta is set substantially above the minimum value that represents material impact on competition for a particular measure. Then the ALECs will face greater risk of a Type II error in the face of disparity constituting material impact than BellSouth would face of at Type I error under parity. In other words, proper balancing would not occur. This problem would be magnified for large sample sizes, because balancing can produce unconventionally large, negative critical values. For example, with samples sizes of 2,500 and 250 for BellSouth and a ALEC, respectively, a delta equal to 0.50 yields a balancing critical value of –3.77, corresponding to a Type I error probability of 0.00008 (i.e., 1 in 12,000), far below any conventional significance level used in statistical testing. A delta equal to 1.00 would yield a balancing critical value of –7.54, corresponding to a microscopically small Type I error probability. Consequently, compelling statistical evidence of discrimination, e.g., a z score of –6.0, might be ignored. Such an outcome would be justified only if one could be certain that delta had not been set too large. If delta is set too large (e.g., delta greater than 0.25), a floor value for the BCV might then be needed. With a delta of 0.25 or less, as recommended by the ALECs, a floor value should be unnecessary.

Q. WHAT DO ALECS RECOMMEND THAT THIS COMMISSION ORDER CONCERNING THE STATISTICAL METHODOLOGY?

A. There are two things that should be included in the Commission's order. First, ALECs propose that the Modified Z be the statistic used for making parity determinations. Second, ALECs propose that this Commission order the parameter delta value be set no higher than 0.25 for all submeasures.

Q. WHEN THE DELTA VALUE FOR THE BALANCING CRITICAL VALUE IS ESTABLISHED, WILL THE ALECS BE SATISFIED THAT THE RECOMMENDED STATISTICAL METHODOLOGY WILL ACCURATELY EVALUATE BELLSOUTH'S PERFORMANCE?

A. Although no perfect methodology for this purpose can be created, I believe that the methodology proposed by the ALECs will be fair to both sides. We expect to monitor how the methodology works in "production mode", when very large amounts of data are being analyzed. AT&T's statistician will monitor how the methodology works after implementation and will make recommendations for improvements, if necessary.

Q. DOES THAT CONCLUDE YOUR TESTIMONY?

A. Yes.

Local Competition Users Group

Statistical Tests for Local Service Parity

February 6, 1998
Membership: AT&T, Sprint, MCI, LCI, WorldCom

Version 1.0

Executive Summary

The Local Competition Users Group has drafted 27 Service Quality Measurements (SQMs) that will be used to measure parity of service provided by incumbent local exchange carriers (ILECs) to competitive local exchange carriers (CLECs). This set of measures includes means, proportions, and rates of various indicators of service quality. This document proposes statistical tests that are appropriate for determining if parity is being provided with respect to these measurements.

Each month, a specified report of the 27 SQMs will be provided by the ILEC, broken down by the requested reporting dimensions. The SQMs are to be systematically developed and provided by the ILECs as specified. Test parameters will be calculated so that the overall probability of declaring the ILEC to be out of parity purely by chance is very small. For each SQM and reporting dimension reported, the difference between the ILEC and CLEC results is converted to a *z*-value. Non-parity is determined if a *z*-value exceeds a selected critical value.

Introduction

Purpose

The Local Competition Users Group (LCUG) is a cooperative effort of AT&T, MCI, Sprint, LCI and WorldCom for establishing standards for the entry of new companies (competitive local exchange carriers, or CLECs) into the local telecommunications market. A key initiative of the LCUG is to establish measures of parity for services provided by incumbent local exchange carriers (ILECs). In short, parity means that the support ILECs provide on behalf of the CLECs is no lesser in quality than the service provided by the ILECs to their own customers.

The LCUG has drafted a document listing service quality measurements (SQMs) that must be reported by the ILECs to insure that CLECs are given parity of support. The SQM document has been submitted to the FCC and made available to PUCs in all 50 states and is pending approval by many of these regulatory agencies. This document has been drafted to describe statistical methodology for determining if parity exists based on the measurements defined in the SQM document.

Service Quality Measurements

The LCUG has identified 27 service quality measurements for testing parity of service. These are:

Category	ID	Description
Pre-Ordering	PO-1	Average Response Interval for Pre-Ordering Information
Ordering and Provisioning	OP-1	Average Completion Interval
	OP-2	Percent Orders Completed on Time
	OP-3	Percent Order Accuracy
	OP-4	Mean Reject Interval
	OP-5	Mean FOC Interval
	OP-6	Mean Jeopardy Interval
	OP-7	Mean Completion Interval
	OP-8	Percent Jeopardies Returned
	OP-9	Mean Held Order Interval
	OP-10	Percent Orders Held >= 90 Days
	OP-11	Percent Orders Held >= 15 Days
Maintenance and Repair	MR-1	Mean Time to Restore
	MR-2	Repeat Trouble Rate
	MR-3	Trouble Rate

	MR-4	Percentage of Customer Troubles Resolved Within Estimate
General	GE-1	Percent System Availability
	GE-2	Mean Time to Answer Calls
	GE-3	Call Abandonment Rate
Billing	BI-1	Mean Time to Provide Recorded Usage Records
	BI-2	Mean Time to Deliver Invoices
	BI-3	Percent Invoice Accuracy
	BI-4	Percent Usage Accuracy
Operator Services and Directory Assistance	OSDA-1	Mean Time to Answer
Network Performance	NP-1	Network Performance Parity
Interconnect / Unbundled Elements and Combos	IUE-1	Function Availability
	IUE-2	Timeliness of Element Performance

The Service Quality Measurements document describes the importance of each measure as an indicator of service parity. The SQM document also describes reporting dimensions that will be used to break each measure out by like factors (*e.g.*, major service group).

Why We Need to Use Statistical Tests

The Telecommunications Act of 1996 requires that ILECs provide nondiscriminatory support regardless of whether the CLEC elects to employ interconnection, services resale, or unbundled network elements as the market entry method. It is essential that CLECs and regulators be able to determine whether ILECs are meeting these parity and nondiscriminatory obligations. In order to make such a determination, the ILEC's performance for itself must be compared to the ILEC's performance in support of CLEC operations; and the results of this comparison must demonstrate that the CLEC receives no less than equal treatment compared to that the ILEC provides to its own operations. Where a direct comparison to analogous ILEC performance is not possible, the comparative standard is the level of performance that offers an efficient CLEC a meaningful opportunity to compete.

When making the comparison of ILEC results to CLEC results, it is necessary to employ comparative procedures that are based upon generally accepted statistical procedures. It is important to use statistical procedures because all of the ILEC-CLEC processes that will be measured are processes that contain some degree of randomness. Statistical procedures recognize that there is measurement variability, and assist in translating results data into useful decision-making information. A statistical approach allows for measurement variability while controlling the risk of drawing an inappropriate conclusion (*i.e,* a "type 1" or "type 2" error, discussed in the next section).

Basic Concepts and Terms

Populations and Samples

Statistical procedures will permit a determination whether the support that the ILECs provide to CLECs is indistinguishable from the support provided by the ILECs to their own customers. In statistical terms, we will determine whether two "samples", the ILEC sample and the CLEC sample, come from the same "population" of measurements.

The procedures described in this paper are based on the following assumption: *When parity is provided, the ILEC data and CLEC data can both be regarded as samples from a common population of possible outcomes.* In other words, if parity exists, the measured results for a CLEC should not be distinguishable from the measured results for the ILEC, once random variability is taken into account. Figure 1 illustrates this concept. On the right side of the figure are histograms of two samples. In this illustration, the ILEC sample contains 200 observations (data values) and the CLEC sample contains 50. Note that the two histograms are not exactly alike. This is due to sampling variation. The assumption that parity exists implies that both samples were drawn from the same population of values. If it were possible to observe this population completely, the population histogram might appear as shown on the left of the Figure. If the samples were indeed taken from this population, histograms drawn for larger and larger samples would look more and more like the population histogram. Figure 1 shows that even when parity is being provided, there will be differences between the samples due to sampling variability. Statistical tests quantify the differences between the two samples and make proper allowance for sampling variability. They assess the chance that the differences that are observed are due simply to sampling variability, if parity is being provided.

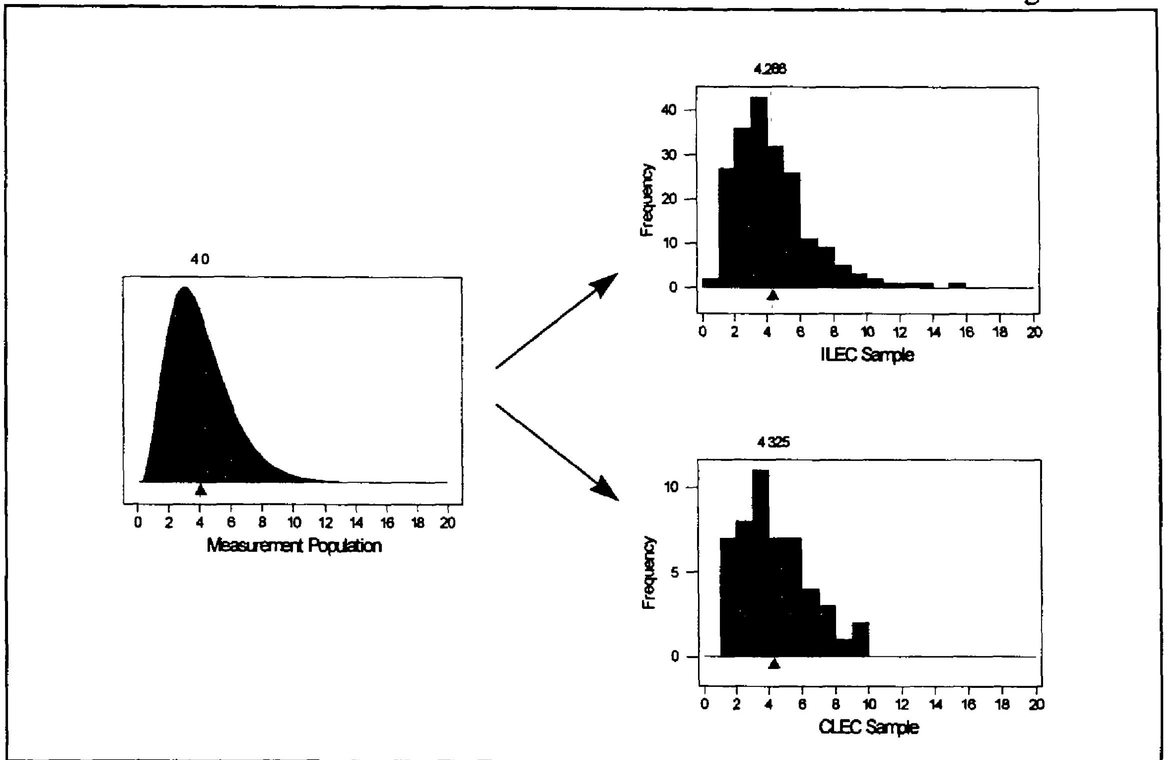

Figure 1.

Measures of Central Tendency and Spread

Often, distributions are summarized using "statistics." For the purpose of this paper, a "statistic" is simply a calculation performed on a sample set of data. Two common types of statistics are known as measures of "central tendency" and "spread."

A measure of central tendency is a summary calculation that describes the middle of the distribution in some way. The most common measure of central tendency is called the "mean" or "average" of the distribution. The mean of a sample is simply the sum of the data values divided by the sample size (number of observations). Algebraically, this calculation is expressed as

$$\bar{x} = \frac{\sum x}{n},$$

where x denotes a value in the sample and n denotes the sample size. The mean describes the center of the distribution in the following way: *If the histogram for a sample were a set of weights stacked on top of a flat board placed on top of a fulcrum (a "see-saw"), the mean would be the position along the board at which the board would balance.* (See Figure 1.) The mean in Figure 1 is indicated by the small triangle at approximately the value "4" on the horizontal axis.

A measure of spread is a summary calculation that describes the amount of variation in a sample. A common measure of spread is a called the "standard deviation" of the sample. The standard deviation is the typical size of a deviation of the observations in the sample from their mean value. The standard deviation is calculated by subtracting the mean value from each observation in the sample, squaring the resulting differences (so that negative and positive differences don't offset), summing the squared differences, dividing the sum by one less than the sample size, then taking the square root of the result. Algebraically, this calculation is expressed as

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}.$$

While the notion of mean and standard deviation exists for populations as well as samples, the mathematical definition for the mean and standard deviation for populations is beyond the scope of this paper. However, their interpretation is generally the same as for samples. In fact, for very large samples, the sample mean and sample standard deviation will be very close to the mean and standard deviation of the population from which the sample was taken.

Sampling Distribution of the Sample Mean

In Figure 1 we showed the positions of the means of the population and the two samples with triangular symbols beneath the distributions. If we sample over successive months, we will get new ILEC samples and new CLEC samples each and every month. These samples will not be exactly like the one for the first month; each will be influenced by sampling variability in a different way. In Figure 2, we show how sets of 100 successive ILEC means and 100 successive CLEC means might appear. The ILEC means can be thought of as being drawn from a population of sample means; this population is called the "sampling distribution" of these ILEC means. This sampling distribution is completely determined by the basic population of measurements that we start with, and the number of observations in each sample. The sampling distribution has the same mean as the population.

Figure 2 illustrates two important statistical concepts:

1. The histogram of successive sample means resembles a bell-shaped curve known as the Normal Distribution. This is true even though the individual observations came from a skewed distribution.

2. The standard deviation of the distribution of sample means is much smaller than the standard deviation of the observations themselves. In fact, statistical theory establishes the fact that the standard deviation on the population of means is smaller by a factor $\sqrt{n}$, where n is the sample size. This effect can be seen in our example: the distribution of the CLEC means is twice as broad as the distribution of the ILEC means, since the ILEC sample size (200) is four times as large as the CLEC sample size (50).

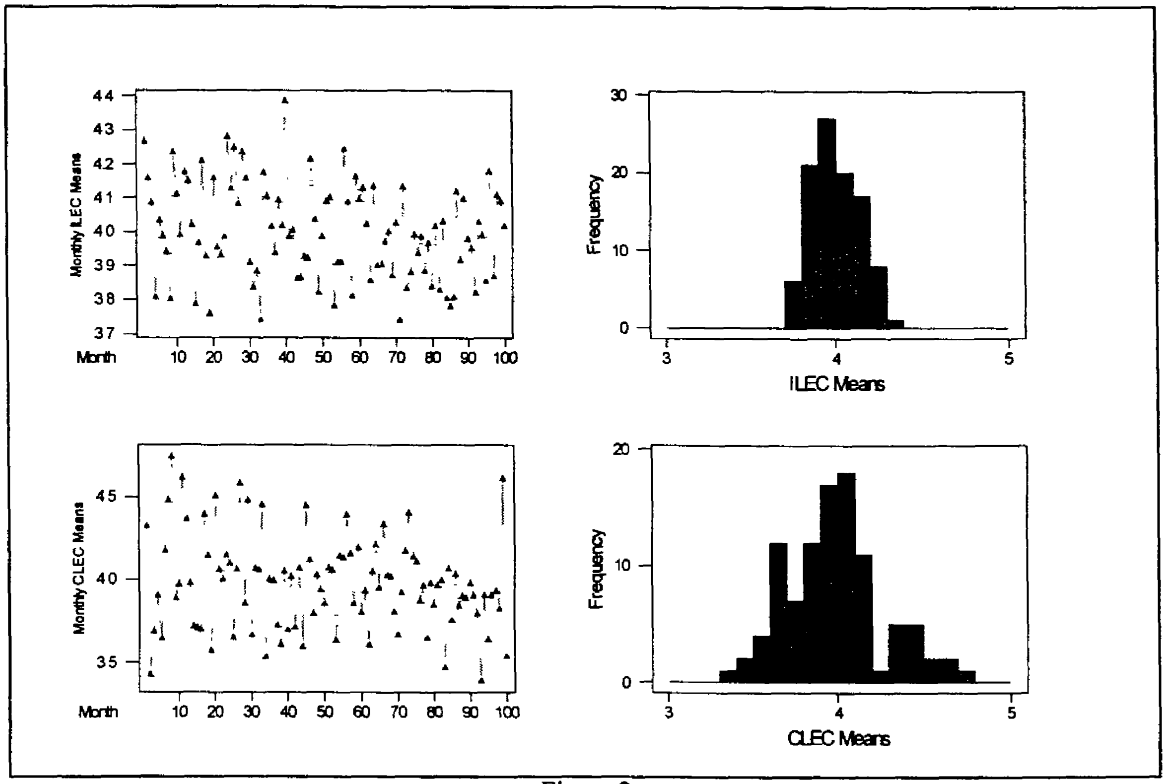

Figure 2.

It is common to call the standard deviation of the sampling distribution of a statistic the "standard error" for the statistic. We shall adopt this convention to avoid confusion between the standard deviation of the individual observations and the standard deviation (standard error) of the statistic. The latter is generally much smaller than the former. In the case of sample means, the standard error of the mean is smaller than the standard deviation of the individual observations by a factor of $\sqrt{n}$.

The Z-test

Our objective is to compare the mean of a sample of ILEC measurements with the mean of a sample of CLEC measurements. Suppose both samples were drawn from the same population; then the difference between these two sample means (*i.e.*, $DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$) will have a sampling distribution which will

(i) have a mean of zero; and
(ii) have a standard error that depends on the population standard deviation and the sizes of the two samples.

Statisticians utilize an index for comparing measurement results for different samples. The index employed is a ratio of the difference in the two sample means (being compared) and the standard deviation estimated for the overall population. This ratio is known as a z-score. The z-score compares the two samples on a standard scale, making proper allowance for the sample sizes.

The computation of the difference in the two sample means is straightforward.

$$DIFF = \bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$$

The standard deviation is less intuitive. Nevertheless, statistical theory establishes the fact that

$$\sigma^2_{\text{DIFF}} = \frac{\sigma^2}{n_{\text{CLEC}}} + \frac{\sigma^2}{n_{\text{ILEC}}},$$

where is the standard deviation of the population from which both samples are drawn. That is, the squared standard error of the difference is the sum of the squared standard errors of the two means being compared.[1]

We do not know the true value of the population because the population cannot be fully observed. However, we can estimate given the standard deviation of the ILEC sample ($_{\text{ILEC}}$).[2] Hence, we may estimate the standard error of the difference with

$$\sigma_{\text{DIFF}} = \sqrt{\frac{\sigma^2_{\text{ILEC}}}{n_{\text{CLEC}}} + \frac{\sigma^2_{\text{ILEC}}}{n_{\text{ILEC}}}} = \sqrt{\sigma^2_{\text{ILEC}}\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

If we then divide the difference between the two sample means by this estimate of the standard deviation of this difference, we get what is called a "*z*-score".

---

[1] Winkler and Hays, *Probability, Inference, and Decision.* (Holt, Rinehart and Winston: New York), p. 370.
[2] Winkler and Hays, *Probability, Inference, and Decision.* (Holt, Rinehart and Winston: New York), p. 338.

$$z = \frac{DIFF}{\sigma_{\mathrm{DIFF}}}$$

Because we assumed that both samples were in fact drawn from the same population, this *z*-score has a sampling distribution that is very nearly Standard Normal, *i.e.*, having a mean of zero and a standard error of one. Thus, the *z*-score will lie between ± 1 in about 68% of cases, will lie between ± 2 in about 95% of cases, and will lie between ± 3 in about 99.7% of cases, always assuming that both samples come from the same population. Therefore, one possible procedure for checking whether both samples come from the same population is to compare the *z*-score with some cut-off value, perhaps +3. For comparisons where the values of z exceed the cutoff value, you reject the assumption of parity as not proven by the measured results. This is an example of a statistical test procedure. It is a formal rule of procedure, where we start with raw data (here two samples, ILEC measurements and CLEC measurements), and arrive at a decision, either "conformity" or" violation".

Type 1 Errors and Type 2 Errors

Each statistical test has two important properties. The first is the probability that the test will determine that a problem exists when in fact there is none. Such a mistaken conclusion is called a type one error. In the case of testing for parity, a type one error is the mistake of charging the ILEC with a parity violation when they may not be acting in a discriminatory manner. The second property is the probability that the test procedure will not identify a parity violation when one does exist. The mistake of not identifying parity violation when the ILEC is providing discriminatory service is called a type two error. A balanced test is, therefore, required.

From the ILEC perspective, the statistical test procedure will be unacceptable if it has a high probability of type one errors. From the CLEC perspective, the test procedure will be unacceptable if it has a high probability of type two errors.

Very many test procedures are available, all having the same probability of type one error. However the probability of a type two error depends on the particular kind of violation that occurs. For small departures from parity, the probability of detecting the violation will be small. However, different test procedures will have different type two error probabilities. Some test procedures will have small type two error when the CLEC mean is larger than the ILEC mean, even if the CLEC standard deviation is the same as the ILEC standard deviation, while other procedures will be sensitive to differences in standard deviation, even if the means are equal. Our proposals below are designed to have small type two error when the CLEC mean exceeds the ILEC mean, whether or not the two variances are equal.

Tests of Proportions and Rates

When our measurements are proportions (*e.g.* percent orders completed on time) rather than measurements on a scale, there are some simplifications. We can think of the "population" as being analogous to an urn filled with balls, each labeled either 0(failure) or 1(success). In this population, the fraction of 1's is some "population proportion". Making an observation corresponds to drawing a single ball from this urn. Each month, the ILEC makes some number of observations, and reports the ratio of failures or successes to the total number of observations; the ILEC does the same does the same for the CLEC. The situation is very similar to that discussed above; however, rather than a wide range of possible result values, we simply have 0's (failures) and 1's (successes). The "sample mean" becomes the "observed proportion", and this will have a sampling distribution just as before. The novelty of the situation is that now the population standard deviation is a known function of the population proportion[3]; if the population proportion is p, the population standard deviation is $\sqrt{p(1-p)}$, with similar simplifications in all the other formulas.

There is a similar simplification when the observations are of rates, *e.g.*, number of troubles per 100 lines. The formulas appear below.

Proposed Test Procedures

Applying the Appropriate Test

Three z-tests will be described in this section: the "Test for Parity in Means", the "Test for Parity in Rates", and the "Test for Parity in Proportions". For each LCUG Service Quality Measurement (SQM), one or more of these parity tests will apply. The following chart is a guide that matches each SQM with the appropriate test.

[illegible]	[illegible]
Preordering Response Interval (PO-1)	**Mean**
Avg. Order Completion Interval (OP-1)	**Mean**
% Orders Completed On Time (OP-2)	**Proportion**
% Order (Provisioning) Accuracy (OP-3)	**Proportion**
Order Reject Interval (OP-4)	**Mean**
Firm Order Confirmation Interval (OP-5)	**Mean**

[3] Winkler and Hays, *Probability, Inference, and Decision.* (Holt, Rinehart and Winston: New York), p. 212.

Mean Jeopardy Interval (OP-6)	**Mean**
Completion Notice Interval (OP-7)	**Mean**
Percent Jeopardies Returned (OP-8)	**Proportion**
Held Order Interval (OP-9)	**Mean**
% Orders Held $\geq$ 90 Days (OP-10)	**Proportion**
% Orders Held $\geq$ 15 Days (OP-11)	**Proportion**
Time To Restore (MR-1)	**Mean**
Repeat Trouble Rate (MR-2)	**Proportion**
Frequency of Troubles (MR-3)	**Rate**
Estimated Time To Restore (MR-4)	**Proportion**
System Availability (GE-1)	**Proportion**
Center Speed of Answer (GE-2)	**Mean**
Call Abandonment Rate (GE-3)	**Proportion**
Mean Time to Deliver Usage Records (BI-1)	**Mean**
Mean Time to Deliver Invoices (BI-2)	**Mean**
Percent Invoice Accuracy (BI-3)	**Proportion**
Percent Usage Accuracy (BI-4)	**Proportion**
OS/DA Speed of Answer (OS/DA-1)	**Mean**
Network Performance (NP-1)	**Mean, Proportion**
Availability of Network Elements (IUE-1)	**Mean, Proportion**
Performance of Network Elements (IUE-2)	**Mean, Proportion**

Test for Parity in Means

Several of the measurements in the LCUG SQM document are averages (*i.e.*, means) of certain process results. The statistical procedure for testing for parity in ILEC and CLEC means is described below:

1. Calculate for each sample the number of measurements (n_{ILEC} and n_{CLEC}), the sample means ($\bar{x}_{\text{ILEC}}$ and $\bar{x}_{\text{CLEC}}$), and the sample standard deviations ($_{\text{ILEC}}$ and $_{\text{CLEC}}$).

2. Calculate the difference between the two sample means; if *larger* CLEC mean indicates possible violation of parity, use *DIFF* = $\bar{x}_{\text{CLEC}} - \bar{x}_{\text{ILEC}}$, otherwise reverse the order of the CLEC mean and the ILEC mean.

3. To determine a suitable scale on which to measure this difference, we use an estimate of the population variance based on the ILEC sample, adjusted for the sized of the two samples: this gives the standard error of the difference between the means as

$$\sigma_{\text{DIFF}} = \sqrt{\sigma^2_{\text{ILEC}}\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
n	mean	variance	n	mean	variance	z	Violation
250	4.038	[illegible]	50	[illegible]	[illegible]	5.15	YES!

Test for Parity in Proportions

Several of the measurements in the LCUG SQM document are proportions derived from certain counts. The statistical procedure for testing for parity in ILEC and CLEC proportions is described below. It is the same as that for means, except that we do not need to estimate the ILEC variance separately.

1. Calculate for each sample sample sizes (n_{ILEC} and n_{CLEC}), and the sample proportions (p_{ILEC} and p_{CLEC}).

2. Calculate the difference between the two sample means; if *larger* CLEC proportion indicates worse performance, use $DIFF = p_{\text{CLEC}} - p_{\text{ILEC}}$, otherwise reverse the order of the ILEC and CLEC proportions.

3. Calculate an estimate of the *standard error for the difference* in the two proportions according to the formula

$$\sigma_{\text{DIFF}} = \sqrt{p_{\text{ILEC}}(1 - p_{\text{ILEC}})\left[\frac{1}{n_{\text{CLEC}}} + \frac{1}{n_{\text{ILEC}}}\right]}$$

4. Hence compute the test statistic

$$z = \frac{DIFF}{\sigma_{\text{DIFF}}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	3.58	Critical value for the test

ILEC			CLEC			Test	
num	den	p	num	den	p	z	Violation
[redacted]	[redacted]	2.00%	[redacted]	[redacted]	17.50%	6.50	YES!

Test for Parity in Rates

A rate is a ratio of two counts, *num/denom*. An example of this is the trouble rate experience for POTS. The procedure for analyzing measurements results that are rates is very similar to that for proportions.

1. Calculate the numerator and the denominator counts for both ILEC and CLEC, and hence the two rates $r_{ILEC} = num_{ILEC}/denom_{ILEC}$ and $r_{CLEC} = num_{CLEC}/denom_{CLEC}$.

2. Calculate the difference between the two sample rates; if *larger* CLEC rate indicates worse performance, use $DIFF = r_{CLEC} - r_{ILEC}$, otherwise take the negative of this.

3. Calculate an estimate of the *standard error for the difference* in the two rates according to the formula

$$\sigma_{DIFF} = \sqrt{r_{ILEC}\left[\frac{1}{denom_{CLEC}} + \frac{1}{denom_{ILEC}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{DIFF}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example:

c:	[redacted]	Critical value for the test

ILEC			CLEC			Test	
num	den	rate	num	den	rate	z	Violation
[redacted]	[redacted]	0.409836	[redacted]	[redacted]	1.133333	6.04	YES!

Exhibit RMB-2

Permutation Analysis Procedural Steps

Permutation analysis is applied to calculate the z-statistic using the following logic:

1. Choose a sufficiently large number T.

2. Pool and mix the CLEC and ILEC data sets

3. Randomly subdivide the pooled data sets into two pools, one the same size as the original CLEC data set (n_{CLEC}) and one reflecting the remaining data points, (which is equal to the size of the original ILEC data set or n_{ILEC}).

4. Compute and store the Z-test score (Z_S) for this sample.

5. Repeat steps 3 and 4 for the remaining T-1 sample pairs to be analyzed. (If the number of possibilities is less than 1 million, include a programmatic check to prevent drawing the same pair of samples more than once).

6. Order the Z_S results computed and stored in step 4 from lowest to highest.

7. Compute the Z-test score for the original two data sets and find its rank in the ordering determined in step 6.

8. Repeat the steps 2-7 ten times and combine the results to determine P = (Summation of ranks in each of the 10 runs divided by 10T)

9. Using a cumulative standard normal distribution table, find the value Z_A such that the probability (or cumulative area under the standard normal curve) is equal to P calculated in step 8.

10. Compare Z_A with the desired critical value as determined from the critical Z table. If $Z_A >$ the designated critical Z-value in the table, then the performance is non-compliant.