国 际 电 信 联 盟

ITU-T E.840

(06/2018)

国际电信联盟
电信标准化部门

E系列：综合网络运行、电话业务、业务运行和人为因素

电信业务质量：概念、模型、指标和可靠性规划 – 电信业务的模型

用于端到端网络性能基准评分和排名的统计框架

ITU-T E.840建议书

国际电信联盟

ITU-T E系列建议书

综合网络运行、电话业务、业务运行和人为因素

国际操作	
定义	E.100-E.103
有关主管部门的一般规定	E.104-E.119
有关用户的一般规定	E.120-E.139
国际电话业务的操作	E.140-E.159
国际电话业务的编号方案	E.160-E.169
国际选路方案	E.170-E.179
用于国内信令系统的信令音	E.180-E.189
国际电话业务的编号方案	E.190-E.199
水上移动业务和公众陆地移动业务	E.200-E.229
国际电话业务中与计费和结算有关的操作规定	
国际电话业务的计费	E.230-E.249
为结算目的对呼叫时长的测量和记录	E.260-E.269
利用国际电话网作非话应用	
概述	E.300-E.319
传真电报	E.320-E.329
有关用户的ISDN规定	E.330-E.349
国际选路方案	E.350-E.399
网络管理	
国际业务统计	E.400-E.404
国际网络管理	E.405-E.419
国际电话业务质量检测	E.420-E.489
业务工程	
话务的测量和记录	E.490-E.505
业务预测	E.506-E.509
确定人工操作的电路数量	E.510-E.519
确定自动和半自动操作的电路数量	E.520-E.539
服务等级	E.540-E.599
定义	E.600-E.649
IP网络的业务工程	E.650-E.699
ISDN业务工程	E.700-E.749
移动网络业务工程	E.750-E.799
电信业务质量：概念、模型、指标和可靠性规划	
与电信业务质量相关的术语和定义	E.800-E.809
电信业务的模型	**E.810-E.844**
电信业务的业务质量指标和相关概念	E.845-E.859
业务质量指标在电信网络规划设计中的使用	E.860-E.879
设备、网络和业务的性能的现场数据收集和评估	E.880-E.899
其他	E.900-E.999
国际操作	
国际电话业务的编号方案	E.1100-E.1199
网络管理	
国际网络管理	E.4100-E.4199

如果需要进一步了解细目，请查阅ITU-T建议书清单。

ITU-T E.840建议书

用于端到端网络性能
基准评分和排名的统计框架

摘要

ITU-T E.840建议书是一系列涵盖端到端网络性能基准测试的建议书中的第一份。ITU-T E.840建议书为网络和服务的性能基准测试提供了一个统计分析框架。该框架描述了用于端到端关键性能指标（KPI）或关键质量指标（KQI）排名的基准方案、使用案例、程序以及统计技术。ITU-T E.840建议书参考了在驾驶或步行测试中使用移动代理（设备）开展的移动业务和基准测试活动，以及利用固定代理或设备在固定位置（例如，商场、办公楼、体育场内）进行的测试。

历史沿革

版本	建议书	批准日期	研究组	唯一ID*
1.0	ITU-T E.840	2018-06-13	12	11.1002/1000/13621

关键词

端到端性能、网络性能基准测试和排名、统计框架。

* 为访问本建议书，请在万维网浏览器的地址栏中输入URL：http://handle.itu.int/，并后跟本建议书的唯一ID。例如：http://handle.itu.int/11.1002/1000/11830-en。

前言

国际电信联盟（ITU）是从事电信领域工作的联合国专门机构。ITU-T（国际电信联盟电信标准化部门）是国际电信联盟的常设机构，负责研究技术、操作和资费问题，并且为在世界范围内实现电信标准化，发表有关上述研究的建议书。

每四年一届的世界电信标准化全会（WTSA）确定ITU-T各研究组的研究课题，再由各研究组制定有关这些课题的建议书。

WTSA第1号决议规定了批准ITU-T建议书须遵循的程序。

属ITU-T研究范围的某些信息技术领域的必要标准，是与国际标准化组织（ISO）和国际电工技术委员会（IEC）合作制定的。

注

本建议书为简明扼要起见而使用的“主管部门”一词，既指电信主管部门，又指经认可的运营机构。

遵守本建议书的规定是以自愿为基础的，但建议书可能包含某些强制性条款（以确保例如互操作性或适用性等），只有满足所有强制性条款的规定，才能达到遵守建议书的目的。“须”或“必须”等其他一些强制性用语及其否定形式被用于表达特定要求。使用此类用语不表示要求任何一方遵守本建议书。

知识产权

国际电联提请注意：本建议书的应用或实施可能涉及使用已申报的知识产权。国际电联对无论是其成员还是建议书制定程序之外的其他机构提出的有关已申报的知识产权的证据、有效性或适用性不表示意见。

至本建议书批准之日止，国际电联尚未收到实施本建议书可能需要的受专利保护的知识产权的通知。但需要提醒实施者注意的是，这可能并非最新信息，因此特大力提倡他们通过下列网址查询电信标准化局（TSB）的专利数据库：http://www.itu.int/ITU-T/ipr/。

目录

页码

ITU-T E.840建议书

用于端到端网络
性能基准评分和排名的统计框架

1 范围

本建议书规定了一个统计框架以及可应用本建议书的基准方案和条件。运营商和监管机构在限定和量化影响用户体验的端到端关键性能指标（KPI）或关键质量指标（KQI）之间的性能差异时需要使用本建议书。

之所以产生对于本建议书的需求，是因为在满足存量用户日益严苛的需要的同时以最优成本扩大用户群的激烈竞争中，运营商对于网络性能的改善已经到了不同运营商之间的差异已变得越来越小的程度。

2 参考文献

下列ITU-T建议书和其他参考文献的条款，通过在本建议书中的引用而构成本建议书的条款。在出版时，所指出的版本是有效的。所有的建议书和其他参考文献都面临修订，使用本建议书的各方应探讨使用下列建议书和其他参考文献最新版本的可能性。当前有效的ITU-T建议书清单定期出版。本建议书引用的文件自成一体时不具备建议书的地位。

[ITU–T E.800] ITU-T E.800建议书（2008年），*有关服务质量的术语定义。*

[ITU–T E.804] ITU-T E.804建议书（2014年），*移动网络流行业务的关键性能指标问题。*

3 定义

无。

4 缩写词和首字母缩略语

本建议书使用下列缩写词和首字母缩略语：

KPI 关键性能指标

KQI 关键质量指标

QoE 体验质量

MOS 平均意见得分

RF 射频

TCP 传输控制协议

5 惯例

5.2.1 StatScore：表示统计得分，即各网络或运营商相对最佳性能网络的相对整体质量。StatScore按每个业务计算。

5.2.2 GlobalNetScore：表示全球网络得分，即各网络或运营商相对最佳性能网络的相对整体质量。GlobalNetScore按所有业务计算。

5.2.3 StatDiff：表示两个进行比较的关键性能指标（KPI）或关键质量指标（KQI）之间的统计上的显著差异。

5.2.4 THrelv：表示两个特定关键性能指标（KPI）或关键质量指标（KQI）值之间的最小差异，与一项业务的用户相关，高于该值则统计显著性优先。

6 基准方案

网络基准测试通常有两个主要使用案例：内部基准测试和竞争性基准测试。内部基准测试的关注重点是需要对网络的初始部署、开发过程以及新业务和新设备的发布进行评估的持续的、具有成本效益的网络性能的保证和改进。内部基准测试也被用于建立已久和成熟的网络中。此外，在评估活动中，需要考虑建有高速公路和城市的区域，以及感兴趣的研究区域（例如工作场所、商场、体育馆和住宅楼宇）。由运营商自己（或代表运营商的业务公司）进行，以及监管机构为检验竞争和自我排名而进行的竞争性基准测试通常跨区域、交通路径（高速公路、铁路）和城市使用，在跨国运营商集团和成熟网络的案例中甚至跨国使用。

图1总结了这些使用案例、建议的工具和技术类型。经常使用步行测试的方式对区域（例如商场、体育馆和工作场所，通常是室内）进行基准测试。除了传统的路径驾驶或者步行测试活动之外，内部基准测试和某些程度的竞争性基准测试，以及室内方案均可获益于固定的探头型工具。后者具有快速、远程可扩展性和设备独立性的优势。因此，这些工具非常适合室内测试方案和在感兴趣的研究区域推出的新业务以及城市（在某种程度上）使用。此外，可以看到，可以采用后验（*posteriori*）或者先验（*a-priori*）分析技术。基准测试中使用最多的是第一种方法：采集数据，并使用统计显著性来对端到端KPI或KQI性能进行评估和排名；在默认情况下，测量准确度被嵌入统计显著性水平中。先验技术涉及为达到具体统计显著性和测量准确度而对所需的测试探头数量的预计算。该技术通常在测试探头昂贵或测试时间有限的情况下使用。

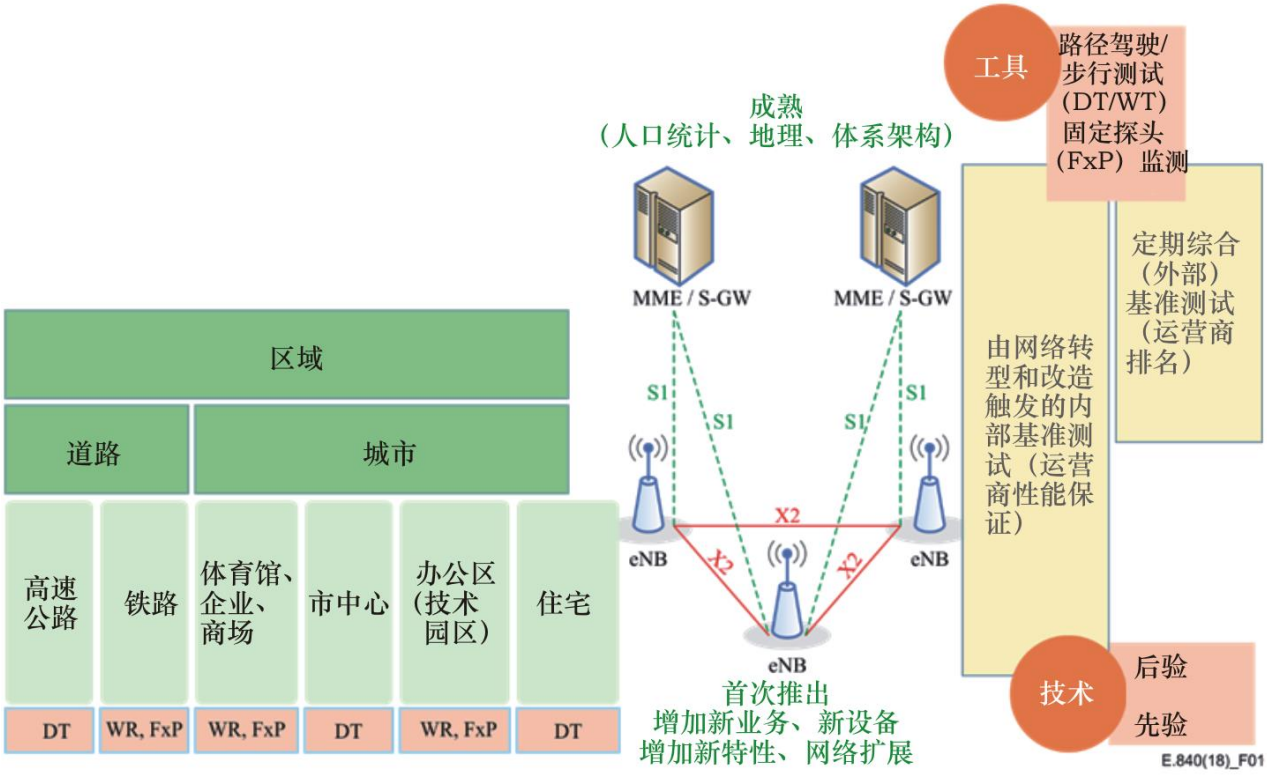

图1 – 基准测试使用案例、建议的工具和技术类型

7 基准测试条件

无论哪种使用案例，基准测试框架需要依靠一套确保一致性、有效性、可靠性和可重复性的先决条件。表1为每个基准测试阶段（设备搭建、测试配置、数据收集、数据处理和分析）提供了所需的先决条件。应注意，表1参考的是确保受到完全控制的测试环境和有效统计分析的最低要求的先决条件。其他ITU-T建议书（如[ITU-T E.804]）提供了测量的具体详细信息。

表1 – 最低要求先决条件指导

基准测试	基准测试先决条件
设备搭建	设备搭建要保持跨网络、平台和设备的一致性；使用相同设备型号用于竞争比较性（“同比”）基准测试
	设备在测试设备供应商规定的适当条件下运行（例如，避免过热，因为可能对设备性能造成不利影响）
测试设置和配置	测试设置要反映真实用户体验；指导见[b-ETSI TR 102581]
	测试设置要避免可能人工影响网络性能的假象：例如：数据服务器位置和传输控制协议（TCP）参数的设置须被验证，以确保所有进行比较的运营商都能获得良好通量。该测试须在下一步数据收集之前进行
	反映用户行为（例如TCP的使用，不同文件、视频或语音通话长度）的不同方案的测试设置脚本处理，并使侵扰最小化，从而不会人为造成网络超载

表1 – 最低要求先决条件指导

基准测试	基准测试先决条件
数据收集	收集测量数据要反映用户体验[例如，每个业务的平均意见得分（MOS），以及对其造成影响的主要网络端到端KPI或KQI]。测量须基于适当的度量标准或者测量工具（根据供应商或者相关建议书提供的指南）
	收集不同时间窗口（高峰或非高峰时间、周末或非周末、假期或非假期）的各类地理或人口统计条件下的测量数据
数据处理和分析	比较收集到的相同设备在相同地区和相同时间窗口期间内的数据 – “同比”比较
	使用统计显著性来进行有意义的比较
	对每个KPI或KQI进行分析

8 基准业务

传统基准移动业务及其KPI或KQI清单以及它们的触发点在本建议书范围之外。更多详细信息见[ITU-T E.800]和[ITU-T E.804]。

如果移动基准测试的范围是要对每个业务进行详细的比较分析，通常在内部基准测试使用案例中进行任务（例如，新设备、新技术扩展等方案），因此建议使用一套综合的端到端KPI和KQI进行分析（在此，KQI是使用质量评估模型（例如语音使用[b-ITU-T P.863]，或视频流使用[b-ITU-T P.1203]）获得的测量值）。此外，建议基于这套KPI和KQI来分析造成每个可能的性能不佳的主要根本原因。

另一方面，如果移动基准测试的目的是进行端到端KPI或KQI性能排名，则通常在比较基准测试使用案例，以及一些内部基准测试方案（例如市场比较、定期市场性能评估）中开展任务，因此可以考虑对每个业务和所有基准业务使用对体验质量产生影响的（QoE；见[b-ITU-T P.10/G.100]）的一套较小的KPI或KQI来进行评分和排名。

本建议书将后一种情况称为竞争性基准测试。其他基准测试系列建议书涵盖此类KPI或KQI集合的详细信息。

9 统计框架

建议的框架旨在从用户角度对网络端到端性能进行评分和排名，可被用于竞争性和内部基准测试。框架定义了数据验证、统计评估度量和显著性检验的流程，以及排名和评分的一般准则。

9.1 数据清洗

为确保基准测试结果有意义并且准确，需要对用作分析输入的数据进行验证。验证的主要内容是数据清洗，意为移除任何测量特有的人工假象和不完整的数据。建议使用最新收集到的数据替换缺失数据。如果5%或以上的数据包含人工假象或缺失，则建议重新收集数据。测量特有人工假象包括但不限于：语音或视频会话服务中的无声通话、在视频流业务中出现数据或视频服务器不可用（例如服务器宕机）时间、持续出现的意料之外的语音、视频

通话或视频流的极低MOS值。在这里，并非由网络本身而是由测试设备或测试仪器造成的任何类型的衰减都应被视为测量特有人工假象。

基准测试数据需要过滤掉这些人工假象，从而确保应用这些统计模型做出的假设的有效性。

9.2 测量统计分布

一般来说，任何测量的KPI或KQI的值的统计分布可以通过一个高斯分布，基于中心极限定理[b-Shaum]来进行粗略估计；通常，样本数目越大，高斯分布的近似值就越准确。建议对所有KPI或KQI值的统计分布进行验证。可以采用两种方式进行。一种可能方式是生成分析的KPI或KQI的测量结果分布图，并通过观察法来验证其正态性。另一种可能方式是使用拟合优度检验来进行正态性验证，例如Kolmogorov-Smirnov（KS）检验、Anderson-Darling（AD）检验或Shapiro-Wilk（SW）检验[b-Mehta]。此外，在罕见或极端的非高斯分布的案例中，可采用非参数检验。ITU-T基准测试系列的其他建议书对各类KPI或KQI实验分布的测试、验证分布正态性的测试和非高斯分布的特别案例进行了讨论。

9.3 基准测试结果的统计性能度量、标准误差和统计显著性

9.3.1 统计性能度量

基准测试分析须以反映平均网络性能（由平均值m表示）或其一致性（由要高于预先定义的阈值的概率Pth表示）的数据性能度量指标为基础。本建议书参考平均统计性能度量指标作为示例。类似的技术可被用于一致性度量。

9.3.2 标准误差

假设测量的KPI或KQI为高斯分布，计算95%置信水平的平均值或Pth的标准误差（见第9.2条）。

因此，取决于KPI或KQI类型、连续分数（例如射频（RF）参数）或MOS或离散分数（例如成败比（r）），根据以下公式给出95%置信水平的标准误差：

StdError(m)=z_95% *std/sqrt(N)= 1.96*std/sqrt(N)

StdError(r)= z_95% * sqrt (r*(1-r)/N) = 1.96 * sqrt (r*(1-r)/N)

如果可用样本不到30个，则高斯分位数z_95%须被替换为Student t_95% (N-1)，表列值，其中N表示可用样本数量。

注 – 标准误差表示测量准确度。因此，若需要达到一个特定准确度，并已知标准偏差的估值，则可以根据第2段中的方程确定在所选置信水平上满足该准确度所需的最少样本数量。这可被用于先验技术（如图1所示），还可以在[b-ITU-T E.802]中用于最小样本数的计算。

9.3.3 统计显著性

置信区间描述一个给定统计置信水平（通常为95%）的标准误差容限。然而，具有相近值和重叠置信区间的两个进行比较的KPI或KQI在统计学上来看不一定相同。两个KPI或KQI之间的准确比较须基于统计显著性。这可以确保错误地拒绝两个KPI或KQI的值相同（在它们确实相同时）的假设的可能性保持在5%。

根据统计显著性测试，可以得出一个运营商的性能优于另一个运营商的结论（在竞争性基准测试使用案例中），以及确定一个新技术或特性是否能够带来明显改进（在内部基准测试使用案例中）。

除了统计显著性之外，在差异与每个KPI或KQI的测量准确度无关或在测量准确度范围之内的时候，还必须使用KPI或KQI特定的相关差阈。在ITU-T基准测试系列其他建议书中定义了KPI或KQI特定的相关差阈（THrelv）。

表2展示了KPI或KQI比较必须如何进行的示例。为保持本建议书的通用性，KQI1和KQI2被用作比较的度量指标的例子。因此，KQI1和KQI2可以是为某一特定业务所选择的KQI（也可以使用KPI）中的任意两个。为KQI1和KQI2计算平均和标准偏差值。使用可用的测试样本数量并应用假设检验方程式A-1和A-2显示，网络1和2在度量指标KQI1和KQI2方面提供的质量统计上来说是相同的。此外，可以看到，每个KQI（KQI1和KQI2）的差异低于THrelv。网络1和2的性能在第三个度量指标KQI3方面显示出统计差异。然而，该差异(0.02)仍然低于THrelv (0.025)，因此不能得出网络1和2性能不同的结论。

表2 – 统计显著性示例

业务	KQI	网络1			网络2			统计@95%置信水平（CL）(Z>1.96)	StatDiff	THrelv
		平均值	标准	N	平均值	标准	N			
业务1	KQI1	3.27	0.3	287	3.35	0.6	212	1.78	否	0.09
	KQI2	0.02	0.14	12	0.015	0.12	10	0.09	否	0.006
	KQI3	0.93	0.26	69	0.91	0.29	71	2.04	是	0.025

对详细的基准测试结果进行的此类分析可被拓展至各类业务以及每个业务的更大的KPI或KQI集合，如第8条所述。此外，根据统计显著性结果（Z统计@95%置信水平（CL），表2），可对各网络的单个KPI或KQI进行排名，如第9.4条所述。

必须注意，除了统计显著性之外，要声称某一网络或业务配置可被认为是“优于”另一个网络或业务配置，还需要ITU-T基准测试系列中其他建议书中定义的某一KPI或KQI特定的相关差阈定义和测量准确度信息。

9.3.4 结果报告

基准测试统计分析和结果报告必须配有详细的测试方案和用于基准测试的条件的描述；否则，对于结果的解读可能会是错误的，导致无意义。

9.4 端到端KPI或KQI评分和排名

可对基准测试活动中考虑的每个区域进行网络端到端KPI或KQI评分和排名。此外，正如已经在第9.3.3条中提到的，需要依靠统计显著性以区分倾向于越来越频繁地出现在当前运营商网络之间的微小性能差异。如果差异与ITU-T基准测试系列其他建议书中描述的差异无关或在测量准确度之内，则必须比较KPI或KQI特定的相关差阈。

附件B提供了一套KPI或KQI的统计评分和排名方法，表3展示了一个示例。

表3 – 统计上显著的端到端KPI或KQI评分和排名示例

	网络1				网络2				
	KPI/KQI	**std**	**N**	**StatDiff**	**KPI/KQI**	**std**	**N**	**StatDiff**	**THrevl**
KPI1/KQI1	0.95	0.22	87	0.05	0.97	0.17	69	0.00	0.018
KPI2/KQI2	0.93	0.26	87	0.00	0.91	0.29	69	0.23	0.019
KPI3/KQI3	3.89	0.50	2600	0.00	3.56	0.70	2070	17.15	0.31
KPI4/KQI4	105.00	5.00	435	42.67	70.00	15.00	350	0.00	34
KPI5/KQI5	1 200.00	300.00	87	0.00	1 800.00	275.00	69	12.31	596

表3显示了用于“同比”比较（见表1指导）的同个区域同一时间窗口的总值。除了性能值，还要计算标准偏差，同时显示测试样本数目。计算95%置信水平（见附件B）的每个KPI或KQI与最佳性能KPI或KQI（表3中黄色高亮部分）相比的具备统计显著差异的StatDiff（见附件B描述）。StatDiff值越低，值越接近最佳性能KPI或KQI；StatDiff = 0表示最佳性能KPI或KQI。此外，可以注意到，在所有案例中，这两个网络的KQI之间的差异高于THrevl，说明统计显著性分析优先。

基于这一分析，表4提供了KPI或KQI排名。

表4 – KPI或KQI统计排名示例

KPI/KQI	网络1	网络2
KPI1/KQI1	排名2	排名1
KPI2/KQI2	排名1	排名2
KPI3/KQI3	排名1	排名2
KPI4/KQI4	排名2	排名1
KPI5/KQI5	排名1	排名2

这一排名可被拓展至更大的KPI或KQI集合，可被考虑用于详细的基准测试以及内部基准测试使用案例。

在一些基准测试使用案例中，可以要求对每项业务以及所有业务的网络统计评分。附件A提供了执行这一要求的可能技术方法。然而，须注意，这类技术方法只有在充分描述和基于技术上支持的假设和条件下才有效。

附件A

应用于移动网络基准测试分析的统计显著性

（此附件是本建议书不可分割的组成部分。）

基准测试分析指的是对描述各个运营商网络性能的KPI或KQI进行的比较。有意义的比较须依靠取决于比较的KPI或KQI的类型、连续性（例如，MOS、射频参数）和比例（例如，完成率或失败率）的统计显著性检验（假设检验）。

在第一种情况中，方程A-1[ITU-T P.1401]决定了显著性差异：

$$Z = StatDiff/sqrt(std1^2/N1 + std2^2/N2) > Zth \quad (A\text{-}1)$$

其中，StatDiff表示进行比较的度量指标之间的差异，std1和std2为标准偏差，N1和N2是每个度量指标进行比较时使用的样本总数。换句话说，如果Z高于Zth（基于超过30个样本的高斯分布，置信水平为CL%），那么StatDiff是置信水平为CL%的统计显著性差异。

在第二种情况中，KPI或KQI比例类型由总样本数中成功或失败的数目p来描述。显著性差异由方程A-2 [b-ITU-T P.1401]计算：

$$Z= StatDiff/sqrt(p1*(1-p1)/N1 + p2*(1-p2)/N2) > Zth \quad (A\text{-}2)$$

其中，p1和p2表示每个进行比较的度量指标的成功或失败数量。

表A.1提供了显著性阈值Zth在不同置信水平的映射。

表A.1 – 不同置信水平的显著性阈值映射

CL%	90	95	96	97	98	99
Zth	1.64	1.96	2.05	2.17	2.33	2.58

如果可用样本少于30个，则须使用t-Student分布，在t-Student (n)中，n = N-1是自由度数目，检验样本总数为N。

须注意，除了统计显著性之外，在差异与每个KPI或KQI的测量准确度无关或在测量准确度范围之内可能出现的时候，还必须使用KPI或KQI特定的相关差阈。在ITU-T基准测试系列其他建议书中定义了KPI或KQI特定的相关差阈（THrelv）。

附件B

网络性能统计评分和排名

（此附件是本建议书不可分割的组成部分。）

本附件描述了用于对在表3的计算中使用的网络的端到端性能进行评分和排名的算法。

– 计算每个网络或运营商的被分析业务的端到端KPI（也可以使用KQI）：

  • KPI_1...KPI_i…KPI_N, i=1，n可以是平均值或中位数或占比（比例）。

– 建立一个基准测试矩阵，j=1，M个网络（运营商），每个业务由N个KPI或KQI度量指标描述 – 见表B.1。

表B.1 – 基准测试矩阵

	Netwk_1……	**Netwk_j…….**	**Netwk_M**
KPI_1…	KPI_1,1	KPI_1,j	KPI_1,M
KPI_i..	KPI_i,1	KPI_i,j	KPI_i,M
KPI_N	KPI_N,1	KPI_N,j	KPI_N,M

– 为矩阵中的每个KPI_i,j计算统计显著性距离。

对于i=1,N

  • 选择最佳值KPI_i,best，“最佳（best）”为拥有最佳网络的网络j

  • 取决于度量指标类型，基于方程B-1和B-2，计算每个KPI_i,j相对最佳值KPI_i,best的统计显著性差异StatDiff_i,j

StatDiff_i,j = max {0,(KPI_i,best-KPI_i,j)/sqrt (std1^2/N1 +std2^2/N2) – Zth} (B-1)

StatDiff_i,j = max {0,(KPI_i,best-KPI_i,j)/sqrt (p1*(1-p1)/N1 + p2*(1-p2)/N2) – Zth} (B-2)

注 – Zth为F(0.05, N1, N2)，95%显著性的统计结果，自由度为N1和N2。

结束

– 为区域的测试业务确定排名第1（Rank 1）的网络或运营商（“最佳性能”）。

排名第1（Rank 1）指的是在所有网络或运营商中，被测业务的统计显著性质量距离最小的网络。

对于 j=1,M

Rank 1 = Rank(j) if StatDiffQuality_min=min(j=1,M) {SUM (i=1,N) {StatDiff_i,j * w_i}}

其中，StatDiff_i,j如方程B-1或B-2定义，w_i表示预先设定的每个业务的加权（如果想要使用加权的话）。否则，可以使用平等统一加权。

结束

为基准测试活动中纳入考虑的所有其他网络或运营商确定排名。

对于 j=1,M

如果Dist(j) =max (0,StatDiffQuality/StatDiffQuality_min – Zth) = 0

Rank (j)= Rank 1

（确定与排名第1网络显示相同统计性能的所有网络）

否则

Dist(j)升序排列

Rank (j) =在矢量Dist(j)中的位置

结束

应注意，应用排名必须根据统计显著性，以及在差异与每个KPI或KQI的测量准确度无关或在测量准确度范围之内时，基于KPI或KQI特定的相关差阈THrelv。在ITU-T基准测试系列其他建议书中定义了KPI或KQI特定的相关差阈（THrelv）。

附录I

网络统计评分和排名的一个可行技术

（此附录不是本建议书不可分割的组成部分。）

有时候，能确定一个总体网络统计分数是比较理想的。为此，每个业务类型的网络性能经常被用作基础标准。应报告每个地区的该类型分数，以及跨区域的总分数，如图1所示。

每个业务的统计分数由影响被分析业务的整体端到端质量的所有考虑的端到端KPI_i或KQI_i (i=1,N)度量指标定义。因此，得分可以由每个KQI的StatDiff_i（见附件B）的加权总和对照第9.4条描述的最佳性能来确定。如可用，StatDiff_i值根据相关差异阈值进行校正。最后结果StatScore描述了进行比较的网络与最佳性能网络相比的端到端性能。

$$StatScore = \Sigma(w\_i*StatDiff\_i)$$

在这里，w_i是分配给每个促进业务质量的KPI或KQI度量指标的权重。StatScore越低，性能越好（或者越接近性能最佳网络），相应排名也越好。

表I.1是表3的新版本，在新版本中增加了一些权重示例；由于权重的定义在本建议书范围之外，这些示例仅是资料性的。如果确定所有KPI或KQI在网络性能整体统计分数中具有相同的重要性，则权重可以一致。然而，应注意，即便考虑相同统一加权，如果KPI或KQI的数量有所增减，网络的统计分数可能发生变化，导致不同的统计结果。

因此，本建议书中的网络的统计评分和排名只有在附有详细的描述和选择的动机以及KPI或KQI的选定权重时才有效。如果没有这一透明度，网络的统计评分和排名无效。

在表I.1的示例中，根据给定权重，网络1以2.15的最低评分获得了最佳排名第1。此外，可以注意到，在所有的案例中，两个网络的KPI或KQI之间的差异均高于适用THrevl，意味着统计显著性分析的结果有效。

表I.1 – 统计评分和排名示例

	网络1				**网络2**				
	KPI	std	N	StatDiff	KPI	std	N	StatDiff	THrevl
KPI1/KQI1	0.95	0.22	87	0.05	0.97	0.17	69	0.00	0.018
KPI2/KQI2	0.93	0.26	87	0.00	0.91	0.29	69	0.23	0.019
KPI3/KQI3	3.89	0.50	2600	0.00	3.56	0.70	2070	17.15	0.31
KPI4/KQI4	105.00	5.00	435	42.67	70.00	15.00	350	0.00	34
KPI5/KQI5	1 200.00	300.00	87	0.00	1 800.00	275.00	69	12.31	596
StatScore				2.15				5.83	
排名				1				2	

一个完整的基准测试活动可针对每个接受测试网络所有支持的业务（j=1，M为支持业务数量）的整体统计分数的计算。必须注意，任何此类整体评分必须出于良好动机，各个业务权重必须明确透明，以遵守本建议书要求。

这一评分可以通过将附件B中计算的每个业务所有统计分数相加来计算，如表I.1示例所示。在这种情况下，最佳性能网络的整体统计分数为零。分数越低，相应网络的性能越好，按照每个业务KPI或KQI权重和合计的各自示例所规定。此类操作得出的分数经常用于表示进行比较的网络与最佳性能相比的整体网络排名：

$$GlobalNetScore = \Sigma(Wser\_i*StatScore\_j)$$

GlobalNetScore的权重Wserv_j的定义在本建议书范围之外。不过，本建议书提供了一些指导。

GlobalNetScore可被作为加权总和来计算，与用于StatScore的计算的权重类似，取决于运营商政策或者业务重点。如果考虑进行内部基准测试方案（图1）的话，权重、数量以及哪些业务被纳入GlobalNetScore计算中可以由运营商选择或决定。

或者，可以根据用户统计分析（例如，众包）、使用的业务类型和不同类型区域的使用比例来决定权重。

不过，整体评分的有效性取决于详细描述和KPI或KQI的选择动机，以及每个业务的选定权重和每个业务在整体评分中的权重。没有此透明度，则网络的统计整体评分和排名无效。不仅如此，正如本建议书正文所述，StatDiff分数必须根据不同KPI或KQI的业务间差异的相关性来处理，如果任何差异可能具有显著性，须被设置为0，但从用户角度来看，这对网络的更佳性能没有帮助。

参考书目

[b-ITU-T E.802] ITU-T E.802建议书（2007年），确定和应用QoS参数的框架和方法。

[b-ITU-T P.10] ITU-T P.10/G.100建议书（2017年），性能和服务质量词汇。

[b-ITU-T P.863] ITU-T P.863建议书（2018年），感知客观收听质量预测。

[b-ITU-T P.1203] ITU-T P.1203建议书（2017年），可靠传送之上渐进下载和自适应视听流服务的参数比特流质量评定。

[b-ITU-T P.1401] ITU-T P.1401建议书（2012年），客观质量预测模型的统计评估、资格审查和比较的方法、度量指标和程序。

[b-ETSI TR 102 581] ETSI TR 102 581, V1.2.1 (2015), *Speech processing, transmission and quality aspects (STQ); A study on the minimum additional required attenuation on the antenna path of the field test equipment*. https://www.etsi.org/deliver/etsi_tr/102500_102599/102581/01.02.01_60/tr_102581v010201p.pdf

[b-Mehta] Mehta, S. (2014). *Statistics topics*.CreateSpace. 160 pp.

[b-Shaum] Spiegel, M.R., Schiller, J.J., Srinivasan, R.A. (2013). Schaum's outlines: *Probability and statistics*, 4th edition.New York, NY: McGraw-Hill.424 pp.

ITU-T建议书系列

系列A	ITU-T工作的组织
系列D	资费及结算原则和国际电信/ICT的经济和政策问题
系列E	**综合网络运行、电话业务、业务运行和人为因素**
系列F	非话电信业务
系列G	传输系统和媒质、数字系统和网络
系列H	视听及多媒体系统
系列I	综合业务数字网
系列J	电视、声音节目和其他多媒体信号的有线网络和传输
系列K	干扰的防护
系列L	环境与ICT、气候变化、电子废物、节能；线缆和外部设备的其他组件的建设、安装和保护
系列M	电信管理，包括TMN和网络维护
系列N	维护：国际声音节目和电视传输电路
系列O	测量设备技术规范
系列P	电话传输质量、电话安装及本地线路网络
系列Q	交换和信令；以及相关的测量和测试
系列R	电报传输
系列S	电报业务终端设备
系列T	远程信息处理业务的终端设备
系列U	电报交换
系列V	电话网上的数据通信
系列X	数据网和开放系统通信及安全
系列Y	全球信息基础设施，互联网的协议问题，下一代网络，物联网和智慧城市
系列Z	用于电信系统的语言和通用软件

瑞士印刷
2020年，日内瓦